

# Targeted high-resolution sensing of volatile organic compounds by covalent nanopore detection

Received: 1 February 2025

Accepted: 11 September 2025

Published online: 24 October 2025

Check for updates

Lauren E. McGivern<sup>1,2</sup>, Zhong Hui Lim<sup>1,2</sup>, Yizhi Yuan<sup>1</sup>, Zonghua Bo<sup>1</sup>, Guangqi Wu<sup>1</sup>, Hagan Bayley<sup>1</sup>✉ & Yujia Qing<sup>1</sup>✉

Volatile organic compounds are choice analytes in a variety of contexts. For example, humans release over 4000 volatile organic compounds, many of which are diagnostic of life-threatening medical conditions. The analysis of combinations of a large number of potential analytes requires the application of costly, cumbersome technology. In this work, we show that covalent nanopore sensing can be used for the targeted detection of a reduced set of analytes in a mixture. In this case aldehydes, which constitute ~5% of human volatiles, can be selectively detected by using reversible thiol-aldehyde chemistry. Further, nanopore engineering permits high-resolution detection, which allows closely related aldehydes, including isomers, to be distinguished. Differential sensing of members of other chemical classes, such as mono alcohols, is also demonstrated by leveraging their enzymatic conversion into aldehydes. Our approach is compatible with the use of low-cost, portable, user-friendly diagnostic devices applicable to a wide variety of objectives, including pollutant monitoring, food and beverage testing and the quality control of pharmaceuticals, as well as disease diagnostics.

Aldehydes are ubiquitously produced through chemical processes across various biological, environmental, and industrial contexts. These include lipid peroxidation in the human body<sup>1</sup>, incomplete combustion during wood burning<sup>2</sup>, and Strecker degradation during beer storage<sup>3</sup>. Single-molecule detection of diverse aldehydes against background chemicals therefore offers wide applications.

For example, small-molecule aldehydes are found in human breath or bodily fluids, alongside a spectrum of volatile organic compounds (VOCs) generated through metabolic processes<sup>1</sup> or by the gut microbiome<sup>4</sup>. Metabolic alterations in specific conditions, such as cancers<sup>5</sup>, gastrointestinal diseases<sup>6</sup>, respiratory disorders<sup>7</sup>, and viral infections<sup>8</sup>, lead to distinct VOC profiles, which offer a non-invasive window into the body's internal biochemical processes. The current gold standard for detection of small molecules is liquid or gas chromatography-mass spectrometry (LC/GC-MS)<sup>9</sup>, which generates a

near-complete profile of all collected VOCs, but requires centralised labs that use expensive equipment and sophisticated analysis packages.

The targeted detection of aldehydes, comprising about 170 species among over 4000 VOCs of human origin<sup>10</sup>, presents an appealing approach to reduce the complexity of small-molecule fingerprints, while retaining significant diagnostic value. Aldehyde mixtures in breath or bodily fluids commonly contain linear or branched, unsaturated or saturated carbon chains, containing from 1 to 17 carbon atoms, as well as aromatic species such as benzaldehyde and its derivatives<sup>10</sup>. A ratiometric fingerprint of aldehydes holds promise for disease detection and health monitoring, provided that tests can be made rapid, simple and inexpensive. For example, the ethanal:pentanal:heptanal ratio in breath shifts from 112:1:1.5 in healthy individuals to 24:1.7:1 in patients with lung cancer<sup>11</sup>, and similarly, the hexanal:heptanal ratio in urine changes from 1.3:1 to 2.5:1<sup>12</sup>. For COVID

<sup>1</sup>Department of Chemistry, Chemistry Research Laboratory, University of Oxford, Oxford, UK. <sup>2</sup>These authors contributed equally: Lauren E. McGivern, Zhong Hui Lim. ✉ e-mail: [hagan.bayley@chem.ox.ac.uk](mailto:hagan.bayley@chem.ox.ac.uk); [yujia.qing@chem.ox.ac.uk](mailto:yujia.qing@chem.ox.ac.uk)

infection, the abundance of octanal and benzaldehyde relative to each other is a hallmark of recent infection<sup>8</sup>.

We have established single-molecule covalent sensing mediated by protein nanopores to detect analytes based on chemical reactivity<sup>13</sup>. The concept was first demonstrated with engineered  $\alpha$ -hemolysin ( $\alpha$ HL) pores<sup>14–17</sup> and later validated by others using alternative nanopores<sup>18–20</sup>. Analyte molecules form covalent bonds reversibly or irreversibly with a sensing group on the internal surface of a pore, thereby generating characteristic changes in the ionic current flowing through the pore under a transmembrane potential. When examining mixed analytes, the current signatures (e.g., the current blockade amplitude and the noise during blockades) identify analytes, while the frequency of reversible events reveals the concentration of individual analytes<sup>21</sup>.

One general challenge is to identify sensing chemistry with suitable kinetics to ensure rapid and quantitative detection of analytes; the ideal lifetimes of covalent analyte-pore adducts and the intervals between sensing events are tens to hundreds of milliseconds, which can be reliably measured by electrical recording. This requirement has so far limited the practical application of covalent sensing. Another challenge is the rational engineering of nanopore sensors to resolve structures with minor variations, such as chain isomers that differ by the position of a methyl group. To date, this has largely relied on trial and error. A more systematic investigation is essential to guide the future development of covalent sensors.

In this work, we focus on the development of a single-molecule, real-time detection technology for rapid ratiometric profiling of aldehydes. We exploit hemithioacetal chemistry for the covalent sensing of aldehydes within a thiol-containing  $\alpha$ HL nanopore, achieving rapid, high-resolution analysis by the formation of short-lived adducts that are identified by a machine learning algorithm. The diastereomeric adducts produce current signatures with differences that are accentuated by rational engineering of nanopores, affording information that discriminates between closely related molecular isomers. We extend the scope of our approach to mono alcohols by selectively converting them to aldehydes. Hence, we have developed a versatile means for the targeted detection of a subset of analytes present within a complex mixture of molecules.

## Results

### Covalent detection of aldehydes through reversible hemithioacetal formation

Stochastic sensing of aldehyde analytes was achieved using reversible thiol-aldehyde chemistry. To eliminate background reactions, notably imine formation and metal chelation, four mutations were introduced in wild-type (WT)  $\alpha$ HL (i.e., AG =  $\alpha$ HL-K8A-M113G-K131G-K147G) (Supplementary Fig. 1)<sup>22</sup>. A single-cysteine mutation was then introduced at position 115 (i.e., AG-T115C). Heteroheptameric nanopores containing one cysteine-bearing subunit, (AG)<sub>6</sub>(AG-T115C), were prepared and used for covalent detection (Fig. 1a, see Methods). (AG)<sub>6</sub>(AG-T115C) carried a single-channel current ( $I_p$ ) of  $-129 \pm 3$  pA at  $-50$  mV ( $N = 80$  pores) (recording buffer: 2 M KCl, 200 mM PIPES and 20  $\mu$ M EDTA at pH 6.8).

The introduction of an aldehyde in the trans compartment (Fig. 1a) produced reversible current blockades, which were attributed to the formation of hemithioacetal adducts (Fig. 1b, c and Supplementary Figs. 2–10, see Supplementary Section 2). Residual currents ( $I_{res}$ ) after hemithioacetal formation are given as percentages of the open pore current ( $I_{res\%} = I_{res}/I_p \times 100\%$ ) (Supplementary Table 1). For example, for propanal,  $I_{res\%} = 98.7 \pm 0.1\%$  ( $N = 3$  pores,  $>300$  events). Aldehydes exist in both hydrated and non-hydrated forms in aqueous solution<sup>23</sup>. The rates of hemithioacetal formation ( $\nu_{on}$ ) were expressed in terms of the total aldehyde concentration and were consistent with bimolecular kinetics (i.e., for propanal,  $\nu_{on} = k_{on}[\text{propanal}]_{tot}$ , where  $k_{on}$  is the observed rate constant of adduct formation). The rates of

hemithioacetal dissociation ( $\nu_{off}$ ) were independent of propanal concentration, consistent with a unimolecular step (i.e.,  $\nu_{off} = k_{off}$ , where  $k_{off}$  is the rate constant of adduct dissociation) (Fig. 1d and Supplementary Figs. 2–9). Both the association and dissociation reactions were pH-dependent and single-channel recordings were performed at pH 6.8. At this pH value, frequent events occurred at  $\mu$ M–mM analyte concentrations (e.g.,  $\sim 500$  events within 10 min for 3 mM butanal), and the mean lifetimes of the hemithioacetal adducts were long enough to allow accurate aldehyde identification from  $I_{res\%}$  values (e.g.,  $\sim 130$  ms for butanal).

In total, 10 different aldehydes were characterised with the (AG)<sub>6</sub>(AG-T115C) nanopore (Fig. 2a and Supplementary Table 1), ranging from straight-chain to branched-chain to aromatic aldehydes. We demonstrated single CH<sub>2</sub> resolution in distinguishing straight-chain aldehydes (Fig. 2b and Supplementary Table 1). From ethanal to octanal, each additional CH<sub>2</sub> group reduced the  $I_{res\%}$  value by  $\sim 0.7\%$ . Interestingly, our approach was able to distinguish between diastereomeric hemithioacetal adducts with opposite chirality at the C $\alpha$  position for both heptanal and octanal (i.e.,  $\Delta I_{res\%} = 0.29 \pm 0.03\%$  for heptanal ( $N = 3$  pores,  $>100$  events for each diastereomer);  $0.39 \pm 0.05\%$  for octanal ( $N = 2$  pores,  $>40$  events for each diastereomer)). We arbitrarily assigned adducts with larger  $I_{res\%}$  as diastereomers A and those with smaller  $I_{res\%}$  as diastereomers B (i.e.,  $I_{res\%,A} > I_{res\%,B}$ ). For shorter straight-chain aldehydes, diastereomeric adducts were not clearly separated by (AG)<sub>6</sub>(AG-T115C) with  $\Delta I_{res\%} < 0.2\%$ . Aromatic aldehydes produced higher  $I_{res\%}$  values than saturated aliphatic aldehydes with similar masses (i.e.,  $I_{res\%} = 97.5 \pm 0.1\%$  for benzaldehyde ( $N = 3$  pores,  $>300$  events) and  $96.1 \pm 0.1\%$  for hexanal ( $N = 3$  pores,  $>400$  events);  $I_{res\%} = 96.6 \pm 0.1\%$  for phenylacetaldehyde ( $N = 4$  pores,  $>400$  events) and  $95.7 \pm 0.1\%$  and  $95.4 \pm 0.1\%$  for heptanal ( $N = 3$  pores,  $>100$  events for each diastereomer)). We speculate that the open-chain aldehydes extend away from the protein wall, creating larger steric blockades than the corresponding cyclic aromatic aldehydes. Branched-chain aldehydes and their straight-chain isomers produced similar  $I_{res\%}$  values (i.e.,  $I_{res\%} = 98.1 \pm 0.1\%$  for 2-methylpropanal ( $N = 5$  pores,  $>500$  events) and  $97.9 \pm 0.1\%$  for butanal ( $N = 4$  pores,  $>400$  events)), with  $\Delta I_{res\%} \sim 0.2\%$  for 2-methylpropanal and butanal simultaneously recorded with the same pore ( $N = 1$  pore,  $>600$  events).

### Simultaneous detection of aldehydes facilitates ratiometric profiling

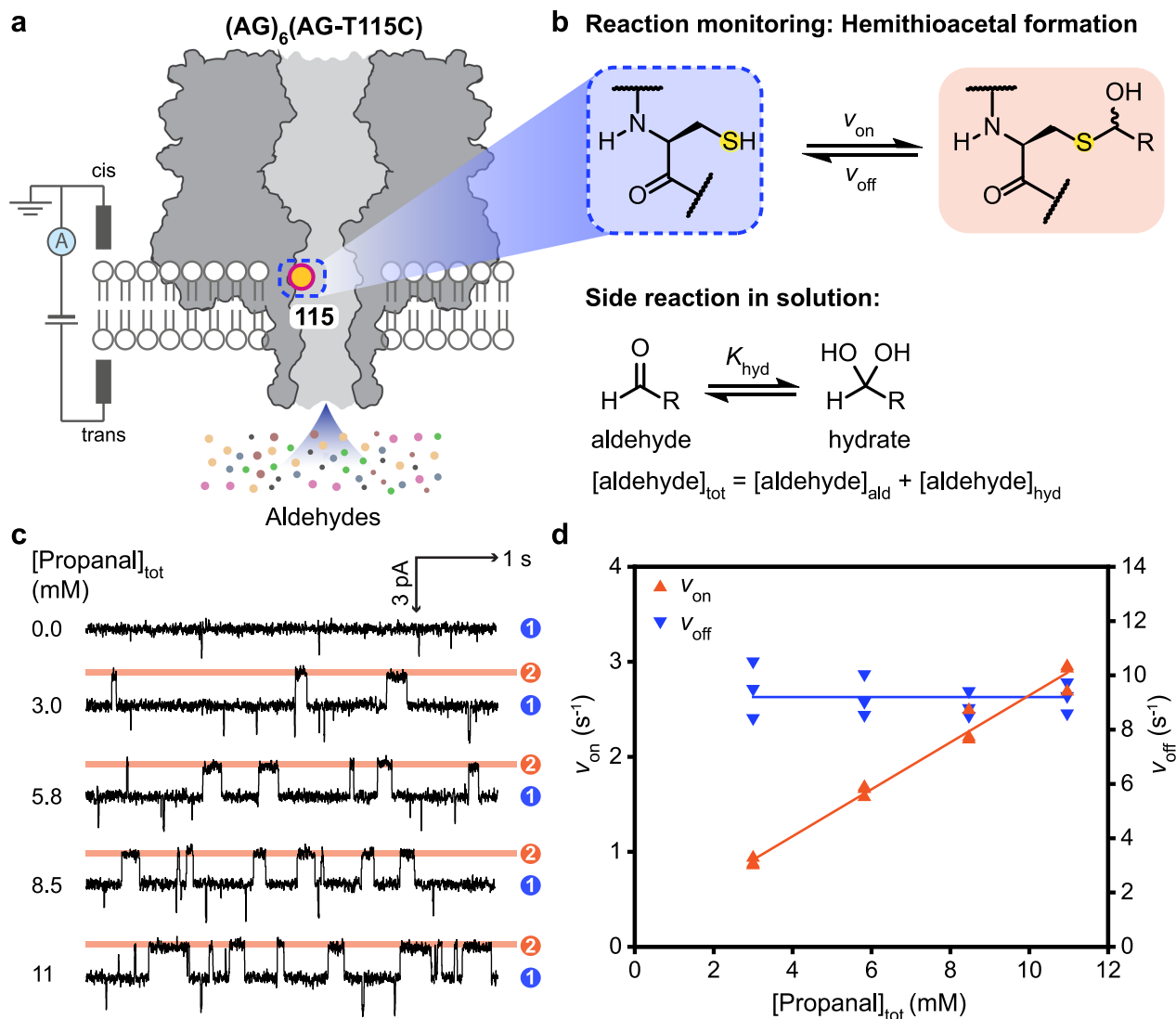
To enable quantitative aldehyde detection, kinetic analysis of the formation and dissociation of hemithioacetal adducts was carried out. The solubility of each aldehyde in the recording solution was measured by NMR in the presence of internal standards (5 mM calcium formate and 5 mM maleic acid). Aldehydes formed hydrates to varying extents in aqueous solutions, and the hydrated forms were unreactive towards thiols and thus not detected by single-channel recording (Fig. 1b)<sup>24</sup>. We determined the hydration equilibrium constants ( $K_{hyd}$ ) by <sup>1</sup>H NMR under electrical recording conditions (Supplementary Tables 2, 3 and Supplementary Fig. 11) and derived the corrected rate constants of adduct formation ( $k_{on}'$ ) by using the equations:

$$\nu_{on} = k_{on}[\text{aldehyde}]_{tot} = k_{on}'[\text{aldehyde}]_{ald} \quad (1)$$

$$K_{hyd} = [\text{aldehyde}]_{hyd}/[\text{aldehyde}]_{ald} = [\text{aldehyde}]_{tot}/[\text{aldehyde}]_{ald} - 1 \quad (2)$$

$$k_{on}' = k_{on} \times (1 + K_{hyd}) \quad (3)$$

$\nu_{on}$ , rate of hemithioacetal formation;  $k_{on}$ , observed rate constant of hemithioacetal formation;  $k_{on}'$ , corrected rate constant of hemithioacetal formation;  $K_{hyd}$ , hydration equilibrium constant;



**Fig. 1 | Single-molecule covalent sensing of aldehydes.** **a** The engineered  $\alpha$ -hemolysin ( $\alpha$ HL) nanopore,  $(AG)_6(AG-T115C)$ , bears a single cysteine at position 115 on one of the seven subunits (boxed). **b** An aldehyde molecule was detected upon reaction with the cysteine thiol to form a hemithioacetal adduct. Aldehydes are hydrated to varying extents in aqueous solutions. Concentrations of free aldehyde were calculated by using the hydration constant  $K_{hyd}$  determined by  $^1H$  NMR. **c** Single-channel recordings at  $-50$  mV (trans) with 0.0, 3.0, 5.8, 8.5 or 11 mM propanal (trans) in 2 M KCl, 200 mM PIPES and 20  $\mu$ M EDTA at pH 6.8. Signals were low-

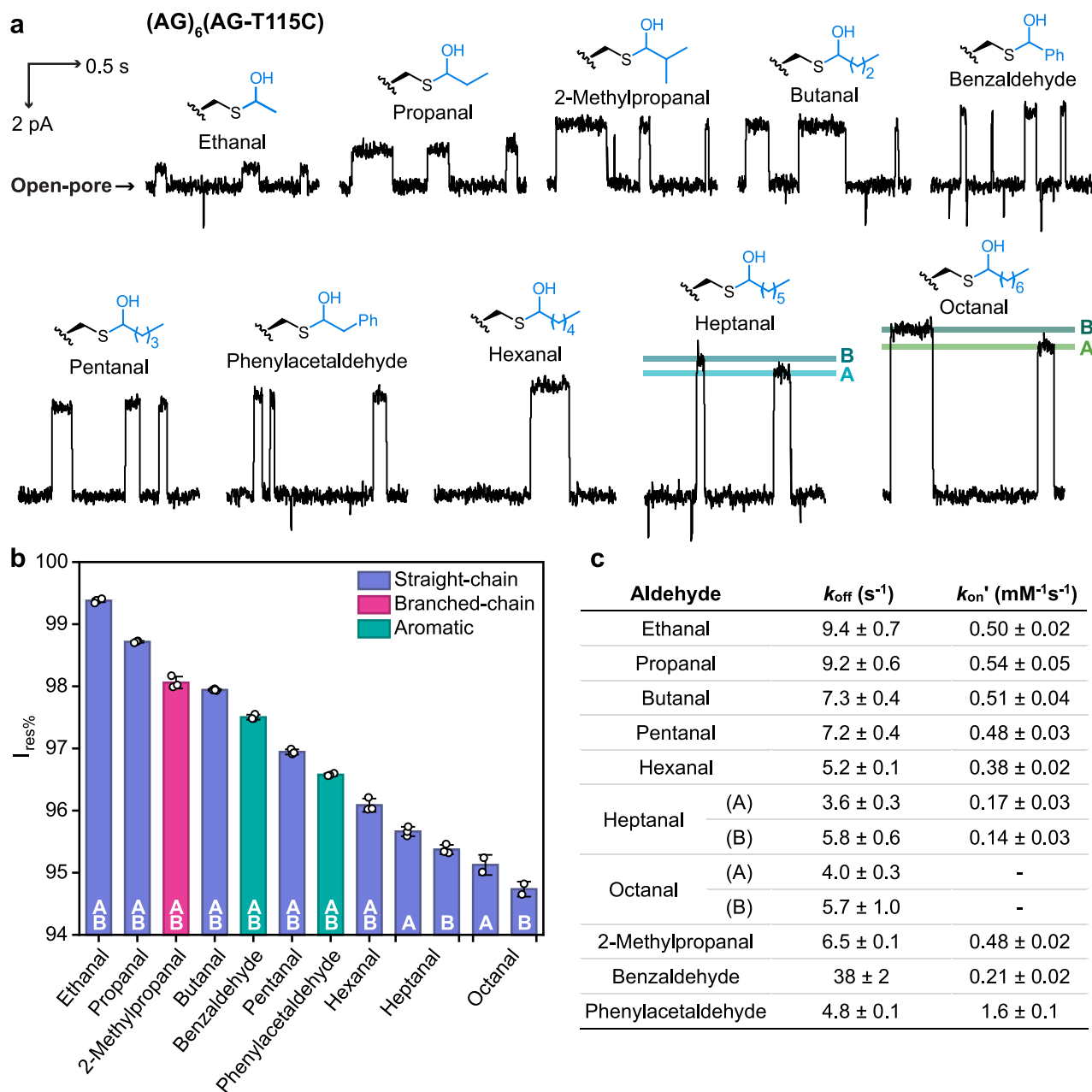
pass filtered at 10 kHz and sampled at 50 kHz. Traces were further filtered at 100 Hz for display. Current levels corresponded to the hemithioacetal adducts formed (level 2, orange) and the unoccupied nanopore (level 1, blue). **d** Rates of adduct formation ( $v_{on}$ ) and dissociation ( $v_{off}$ ), recorded with individual pores, plotted against total propanal concentration showing bimolecular kinetics for the forward association step (level 1 to 2), and unimolecular kinetics for the reverse dissociation step (level 2 to 1). Source data are provided as a Source Data file.

$[aldehyde]_{tot}$ , total concentration of aldehyde;  $[aldehyde]_{hyd}$ , concentration of aldehyde in the hydrate form;  $[aldehyde]_{ald}$ , concentration of free aldehyde.

For the bimolecular formation of hemithioacetal adducts,  $k_{on}$  gradually decreased as chain length increased, from  $0.5 \text{ mM}^{-1}\text{s}^{-1}$  for propanal to  $0.3 \text{ mM}^{-1}\text{s}^{-1}$  for heptanal (Fig. 2c). While the energy difference is small, these observations could reflect the steric hindrance for the thiolate to approach the carbonyl group along the Bürgi-Dunitz angle, caused by the conformationally labile alkyl chains<sup>25</sup>. As rates of adduct dissociation were independent of aldehyde concentration, no correction for hydration was required. For the unimolecular dissociation of hemithioacetal adducts formed with straight-chain aldehydes, values of  $k_{off}$  remained within an order of magnitude, gradually decreasing, as the chain length increased, from  $9.4 \text{ s}^{-1}$  for ethanal to  $\sim 5 \text{ s}^{-1}$  for hexanal, heptanal and octanal (Fig. 2c). This could be attributed to the positive inductive effect of the increasing alkyl chain length.

Simultaneous detection of multiple aldehydes with a single nanopore was demonstrated with a mixture of 7 straight-chain aldehydes (i.e., ethanal to octanal) (Fig. 3a). The consistent current signatures for individual aldehydes and the clear separation between them enabled automated assignment of events by using machine learning (see Supplementary Section 5). Almost 1000 individual events per aldehyde were collected from separate traces of each analyte (Supplementary Fig. 12). A stratified random split was performed, leaving 30% of the events as the test set. Three features were extracted to characterise each event:  $I_{res\%}$ , event duration, and the root-mean-squared noise of the event (see Supplementary Section 5). The random forest model achieved the highest accuracy of 98% on both the training and test sets, which was calculated as the fraction of correctly classified events in the dataset, with manual labels taken as the ground truth (Supplementary Fig. 13).

Additionally, we demonstrated ratiometric measurements of aldehydes in mixtures with pentanal and butanal, for which the detection



**Fig. 2 | Thiol-aldehyde chemistry within the (AG)<sub>6</sub>(AG-T115C) nanopore.**

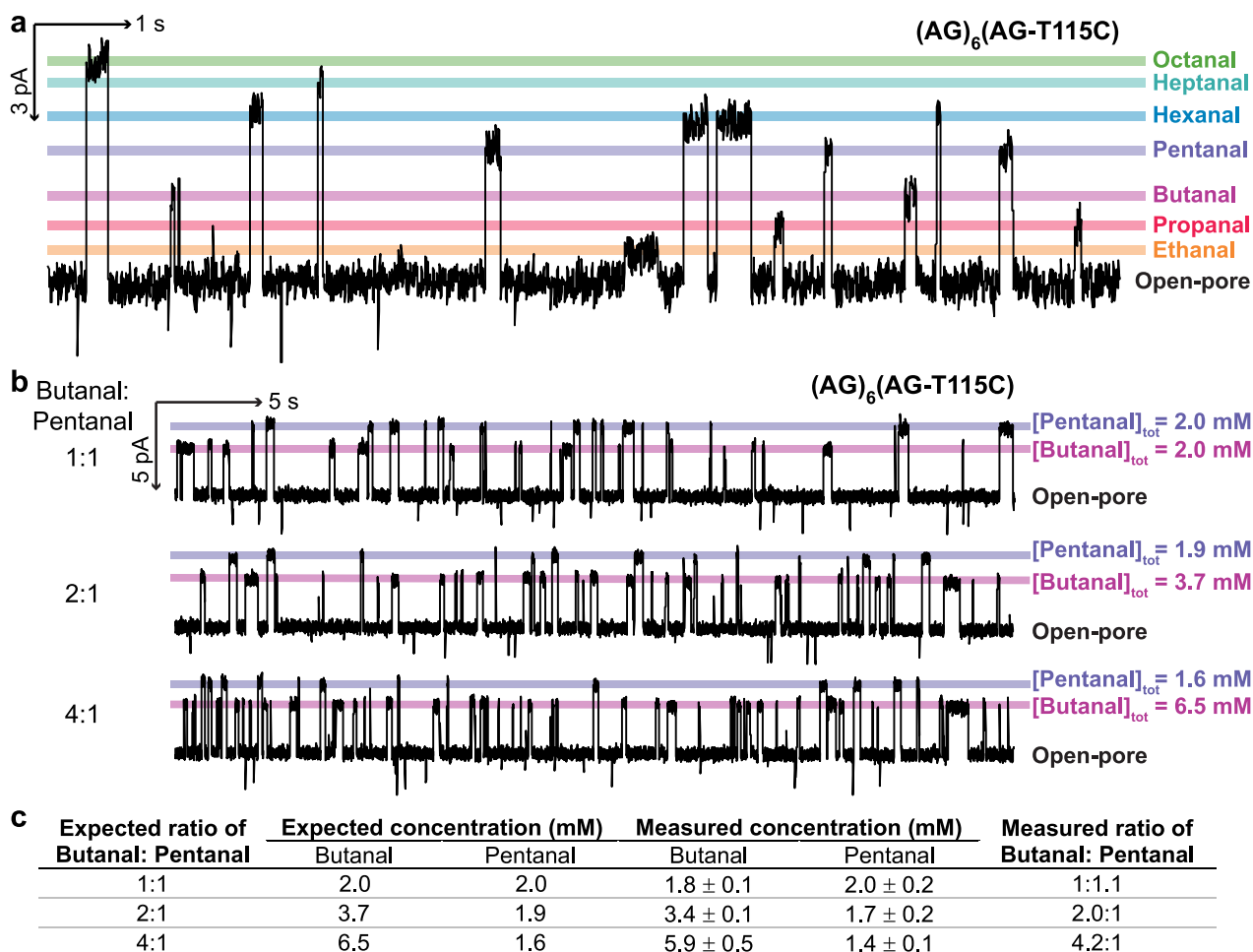
**a** Aldehydes reacted with the cysteine at position 115 to generate hemithioacetal adducts reversibly. Diastereomeric adducts were resolved for heptanal and octanal. Diastereomer A was arbitrarily assigned with a higher  $I_{res}\%$  than diastereomer B. Recording conditions: -50 mV (trans) with 2 M KCl, 200 mM PIPES and 20  $\mu$ M EDTA at pH 6.8. Aldehyde concentrations (trans): 5.8 mM ethanal, 5.8 mM propanal, 6.1 mM 2-methylpropanal, 5.8 mM butanal, 5.7 mM benzaldehyde, 5.5 mM pentanal, 5.5 mM phenylacetaldehyde, 2.7 mM hexanal, 1.5 mM heptanal, 1.3 mM octanal. Signals were low-pass filtered at 10 kHz and sampled at 50 kHz. Traces were further filtered at 100 Hz for display. **b**  $I_{res}\%$  values for the hemithioacetal adducts. Error

bars are standard deviations of  $I_{res}\%$  values determined with >3 different nanopores (Except for octanal which is with 2 different nanopores). **c** Rate constants for adduct formation corrected for hydration ( $k_{on}'$ ) and dissociation ( $k_{off}$ ). Errors are standard deviations across 3 different nanopores (Except for octanal which is for 2 different nanopores). Corrected rate constants for adduct formation were obtained from observed rate constants ( $k_{on}$ ) by using the equation:  $k_{on}' = k_{on} \times (1 + K_{hyd})$ . Rates of adduct formation for octanal were not reported due to the poor solubility of octanal in recording solutions, precluding accurate determinations. Source data are provided as a Source Data file.

frequencies reflected their varied concentration ratios (Fig. 3b, c, see Supplementary Section 6). Within <10 minutes of single-channel recording, ratios of adduct formation events could be used to accurately determine the concentrations of pentanal and butanal (i.e., <10% difference between measured and expected concentrations) (Fig. 3c, see Supplementary Section 6). For disease diagnosis and product quality control applications, a more practical, and equally informative approach, is to directly use the ratios of adduct formation rates as fingerprints of relative concentrations of aldehydes in samples.

### Rational nanopore engineering for diastereomer and structural isomer resolution

Challenged by the similar current signatures observed with aldehyde structural isomers (e.g., butanal and 2-methylpropanal) with the (AG)<sub>6</sub>(AG-T115C) nanopore, rational nanopore engineering was undertaken. We also hypothesised that current level resolution of diastereomeric hemithioacetal adducts would provide an additional layer of information to aid structural isomer resolution. To this end, we designed two additional nanopores in which the reactive cysteine



**Fig. 3 | Single-molecule profiling of aldehyde mixtures.** **a** Simultaneous detection of 7 straight-chain aldehydes at the single-molecule level. Total aldehyde concentrations (trans): 0.7 mM ethanal, 0.7 mM propanal, 0.7 mM butanal, 0.5 mM pentanal, 0.6 mM hexanal, 0.5 mM heptanal, and 0.3 mM octanal. **b** Ratiometric detection of mixtures of butanal and pentanal at varying ratios. Ratios of aldehydes were altered mid-experiment by incrementally replacing the analyte solution in the trans chamber with a butanal stock solution (see Supplementary Section 6).

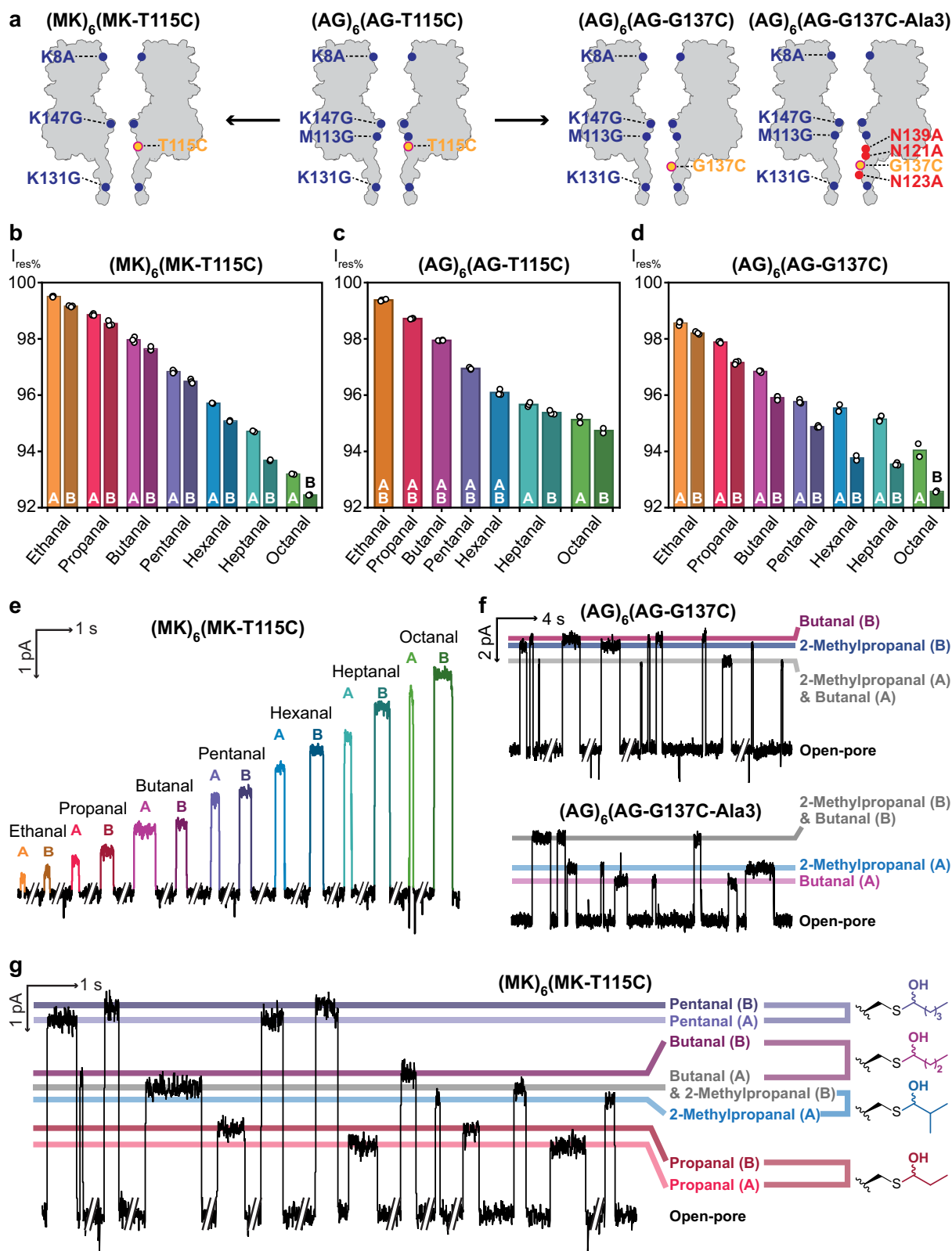
Recording conditions:  $-50$  mV (trans) with 2 M KCl, 200 mM PIPES and 20  $\mu$ M EDTA at pH 6.8. Signals were low-pass filtered at 10 kHz and sampled at 50 kHz. Traces were further filtered at 100 Hz for display. **c** Measured total concentrations of pentanal and butanal were in good agreement with expected total concentrations (i.e.,  $<10\%$  difference). Errors are standard deviations of concentrations determined with 3 different nanopores. Source data are provided as a Source Data file.

residue was positioned within a narrower, and hence more sensitive, region of the nanopore  $\beta$  barrel (Fig. 4a). In the (MK)<sub>6</sub>(MK-T115C) nanopore (MK =  $\alpha$ HL-K8A-K131G-K147G), methionine residues were reintroduced at position I13 to reduce the internal diameter near the sensing site (Supplementary Fig. 14). In the (AG)<sub>6</sub>(AG-G137C) nanopore, the cysteine residue was moved to position 137, which was in close proximity to the narrowest region of the nanopore bearing an AG background (Supplementary Fig. 14). Three asparagine-to-alanine mutations around the sensing site were further introduced to reduce steric hindrance and promote diastereomeric interactions of hemithioacetal adducts with the local protein environment, leading to the (AG)<sub>6</sub>(AG-G137C-Ala3) nanopore (AG-G137C-Ala3 = AG-N121A-N123A-G137C-N139A) (Fig. 4a).

The  $I_{\text{res}\%}$  of 7 different straight-chain aldehydes (i.e., ethanal to octanal) were characterised with the (MK)<sub>6</sub>(MK-T115C) and (AG)<sub>6</sub>(AG-G137C) nanopores (Fig. 4b–d and Supplementary Tables 4, 5). Improved diastereomer discrimination was observed in both nanopores: diastereomeric hemithioacetal adducts from ethanal to octanal were easily separable in both nanopores:  $0.3\% < \Delta I_{\text{res}\%} < 1.0\%$ , in the (MK)<sub>6</sub>(MK-T115C) nanopore (Fig. 4b, e and Supplementary Table 4) and  $0.4\% < \Delta I_{\text{res}\%} < 1.6\%$ , in the (AG)<sub>6</sub>(AG-G137C) nanopore (Fig. 4d and Supplementary Table 5). Good current level separation was achieved

across all diastereomeric adducts in the (MK)<sub>6</sub>(MK-T115C) nanopore (Fig. 4e), whereas some overlap was seen in the (AG)<sub>6</sub>(AG-G137C) nanopore (e.g.,  $I_{\text{res}\%} = 95.9 \pm 0.1\%$  for diastereomer B of butanal and  $95.8 \pm 0.1\%$  for diastereomer A of pentanal).

Current level resolution of diastereomeric hemithioacetal adducts allowed for structural isomer resolution. In the (AG)<sub>6</sub>(AG-G137C) nanopore, butanal and 2-methylpropanal could be distinguished based on diastereomers B (i.e.,  $\Delta I_{\text{res}\%} = 0.25 \pm 0.02\%$  for diastereomers B of butanal and 2-methylpropanal,  $N = 3$  pores,  $>35$  events for each aldehyde) (Fig. 4f). Further, within the (AG)<sub>6</sub>(AG-G137C-Ala3) nanopore, diastereomer pairs were better resolved for both butanal and 2-methylpropanal; distinct separation between diastereomers A enabled clear resolution of these chain isomers (i.e.,  $\Delta I_{\text{res}\%} = 0.42 \pm 0.03\%$  for diastereomers A of butanal and 2-methylpropanal,  $N = 4$  pores,  $>60$  events for each aldehyde) (Fig. 4f). In the (MK)<sub>6</sub>(MK-T115C) nanopore, butanal and 2-methylpropanal could be distinguished by using diastereomer A of 2-methylpropanal and diastereomer B for butanal (i.e.,  $\Delta I_{\text{res}\%} = 0.43 \pm 0.03\%$  for diastereomer B of butanal and diastereomer A of 2-methylpropanal,  $N = 2$  pores,  $>50$  events for each aldehyde). As a proof of concept, simultaneous detection of propanal, butanal, 2-methylpropanal and pentanal was demonstrated in the (MK)<sub>6</sub>(MK-T115C) nanopore (Fig. 4g). As the



current level blockades for pentanal and propanal do not overlap with those of butanal and 2-methylpropanal, all four aldehydes could be differentiated from  $I_{res\%}$  alone.

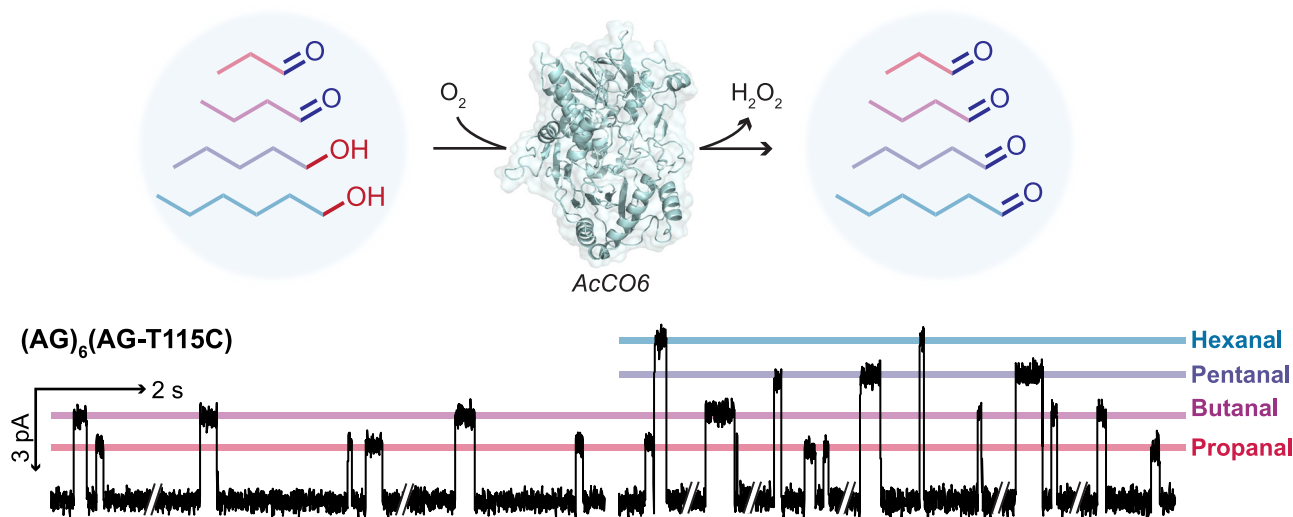
#### Enzyme-assisted differential sensing of alcohols and aldehydes

Nanopore covalent sensing is inherently selective for a targeted class of analytes. For example, in a mixture of alcohols (1-pentanol and 1-

hexanol) and aldehydes (propanal and butanal), only the aldehydes produced events within the (AG)<sub>6</sub>(AG-T115C) nanopore sensor (Fig. 5, Left). After treatment with an engineered alcohol oxidase<sup>26</sup>, additional aldehyde species—pentanal and hexanal—were identified (Fig. 5, Right), revealing the presence of alcohols in the original sample. This simple functional group conversion step, coupled to nanopore sensing, therefore allowed for differential single-molecule detection of

**Fig. 4 | Diastereomer discrimination in engineered nanopores facilitates structural isomer resolution.** **a** Rational engineering of nanopores to enhance diastereomer discrimination through two parallel strategies: Left: narrowing the pore diameter around the sensing group (yellow); Right: moving the sensing group to the narrowest internal site. Mutations are highlighted relative to the WT background in all subunits (blue) or in the cysteine-bearing subunit (red). **b** Within the (MK)<sub>6</sub>(MK-T115C) nanopore, diastereomeric adducts were resolved for all straight-chain aldehydes tested. Diastereomer A was arbitrarily assigned with a higher  $I_{res}$  than diastereomer B. **c** Within the (AG)<sub>6</sub>(AG-T115C) nanopore, diastereomeric adducts were resolved for only heptanal and octanal. **d** Within the (AG)<sub>6</sub>(AG-G137C) nanopore, diastereomeric adducts were resolved for all straight-chain aldehydes

tested. **e** Good current level separation was achieved across all diastereomeric adducts in the (MK)<sub>6</sub>(MK-T115C) nanopore. **f** A pair of chain isomers, butanal and 2-methylpropanal, were distinguished by diastereomers B in the (AG)<sub>6</sub>(AG-G137C) nanopore or by diastereomers A in the (AG)<sub>6</sub>(AG-G137C-Ala3) nanopore. **g** Propanal, butanal, 2-methylpropanal, and pentanal were detected and distinguished within a single (MK)<sub>6</sub>(MK-T115C) pore. Butanal and 2-methylpropanal were distinguished by the diastereomer A of 2-methylpropanal and the diastereomer B of butanal. Recording conditions: -50 mV (trans) with 2 M KCl, 200 mM PIPES and 20  $\mu$ M EDTA at pH 6.8. Signals were low-pass filtered at 10 kHz and sampled at 50 kHz. Traces were further filtered at 50 Hz for display. Source data are provided as a Source Data file.



**Fig. 5 | Enzyme-facilitated indirect alcohol detection.** The (AG)<sub>6</sub>(AG-T115C) nanopore selectively detected aldehydes in an aldehyde-alcohol mixture (Left). After treatment with an engineered alcohol oxidase (AcCO6), additional aldehyde species—pentanal and hexanal—were observed (Right), which indicated the

presence of the corresponding alcohols in the original sample. Recording conditions: -50 mV (trans) with 2 M KCl, 190 mM PIPES and 20  $\mu$ M EDTA at pH 6.8. Signals were low-pass filtered at 10 kHz and sampled at 50 kHz. Traces were further filtered at 100 Hz for display.

alcohols and aldehydes in a mixture—a powerful strategy worth further development.

## Discussion

The present work exploited thiol-aldehyde chemistry, a previously unexplored dynamic covalent chemistry within protein nanopores, to target the sensing of aldehydes, a key class of chemicals ubiquitous in daily life. We achieved single-molecule identification of 10 straight-chain, branched-chain, and aromatic aldehydes, which represent potential biomarkers for lung cancer<sup>5</sup>, Crohn's disease<sup>6</sup>, SARS-CoV-2<sup>8</sup>, as well as being atmospheric pollutants<sup>2</sup>, and beverage impurities<sup>3</sup>. Our approach resolved straight-chain aldehydes that differed by single CH<sub>2</sub> groups, members of hemithioacetal diastereomeric pairs, and structural isomers—challenges often faced by conventional small-molecule detection methods. Compared to traditional LC/GC-MS methods, pre-derivatization of aldehydes (e.g., with hydrazine-based chromophores) was not required<sup>27</sup>. Additionally, we demonstrated rational engineering of the protein environment surrounding the covalent sensing site, achieving improved signal separation for diastereomeric adducts by promoting interactions with the pore interior. Rational engineering of the sensing region thus holds promise for improving small-molecule covalent detection with a given nanopore scaffold.

As a proof of concept, we established ratiometric profiles of mixed aldehydes at mM concentrations with >400 events collected in 10 min with a single nanopore. Given that aldehydic VOCs ranging from propanal to nonanal are exhaled at 3-300 pM concentrations<sup>28</sup>, in

a commercialised device, pre-concentration steps would be coupled to nanopore detection. For example, 3-300  $\mu$ M concentrations of aldehydes can theoretically be obtained by pre-concentrating 5 L of breath volatiles into 5  $\mu$ L of solvent by using solid-phase microextraction (SPME)<sup>29,30</sup>. Detection of analytes in nanolitre droplets (i.e., 300 nL) was previously demonstrated in a microfluidics set-up, and could also be integrated here<sup>31</sup>. Further, selective enrichment of aldehydes has been achieved using SPME devices with fibre coatings selective for aldehydes<sup>28,32,33</sup>—this would simultaneously eliminate potentially interfering chemical classes (e.g., disulfides) that might hinder disease diagnosis. For longer chain aldehydes with poor solubilities (i.e., <1  $\mu$ M solubilities), the concentration of VOCs into solvents such as water/ethanol solutions might be considered, although this would require alternative bilayer systems with greater tolerance to organic solvents. Improved detection frequency might also be achieved by using nanopore engineering, e.g., mutants with multiple covalent sensing sites<sup>14</sup> or mutagenesis of proximate residues to improve cysteine thiol reactivity<sup>34</sup>. In addition, nanopore sensing devices containing arrays of pores would enhance throughput (e.g., the MinION device contains up to 512 active pores per flow cell)<sup>35</sup>. Based on our findings, we envision an accessible nanopore platform for rapid aldehyde detection, paving the way for applications in disease diagnosis, environmental monitoring, and the quality control of food and beverages.

Moving forward, the detection of other biologically relevant chemical classes can be explored by capitalising on the aldehyde-sensing system developed here. As central metabolites, aldehydes are generated by a wide range of enzymes from various functional groups,

including carboxylic acids, primary alcohols, and primary amines<sup>36</sup>. Many of these convertible chemical classes are also potential disease biomarkers. For example, 1-pentanol, detected as 1-pentanal in this work, has been found in the exhaled breath of lung cancer patients but not healthy individuals<sup>5</sup>. While we have demonstrated mono alcohol sensing after enzymatic conversion to aldehydes, the detection of other chemical classes will require the application of means for the direct chemical conversion to reactive molecules or the identification of suitable enzymes for conversion, particularly with respect to substrate scope and catalytic efficiency. In the long term, to leverage thiol-aldehyde sensing chemistry, we envision a versatile sensing workflow that employs a suite of reagents for conversion to aldehydes, followed by rapid single-molecule detection in low-cost, portable, user-friendly devices.

## Methods

### Preparation of plasmids

Construction of plasmids encoding mutant  $\alpha$ HL monomers, together with the primers required, are described in the Supplementary Information. Plasmids are available from the corresponding authors upon request.

### Preparation of nanopore sensors

Nanopore heteroheptamers were prepared according to the procedure below, which is a modification of a method previously reported<sup>37</sup>.

$\alpha$ HL monomers were prepared with an *E. coli* in vitro transcription and translation (IVTT) system (*E. coli* T7 S30 Extract System for Circular DNA, Cat #L1130, Promega). Prior to use, the T7 S30 extract provided in the kit was treated with 1  $\mu$ L rifampicin (1  $\mu$ g/mL, final concentration) to suppress transcription by *E. coli* RNA polymerase. A standard reaction comprised: DNA plasmid mixture (< 4  $\mu$ g, plasmids encoding cysteine-free and cysteine-containing subunits were in an 8:1 ratio), amino acid mixture without methionine (5  $\mu$ L, as supplied in the kit), S30 premix without amino acids (20  $\mu$ L, as supplied in the kit), [<sup>35</sup>S] methionine (2  $\mu$ L, 1200 Ci/mmol, 15 mCi/mL, MP Biomedicals), rabbit red blood cell membranes (2  $\mu$ L, ~1 mg protein/mL), and T7 S30 extract for circular DNA (15  $\mu$ L, as supplied in the kit). The reaction mixture was incubated at 37 °C for 2 h.

Next, MBSA buffer (1 mL; 3-morpholinopropane-1-sulfonic acid (10 mM), NaCl (150 mM), bovine serum albumin (1 mg/mL), pH 7.4) was added to the reaction mixture. The mixture was centrifuged at 25,000  $\times$  g for 10 min at 4 °C. The supernatant was then removed, and the pellet solubilised at room temperature with 2 $\times$  Laemmli sample buffer (25  $\mu$ L) which contains 10% 2-mercaptoethanol. Before loading the samples, a 5.5% SDS/PAGE gel was pre-run with 1 $\times$  Tris-Glycine SDS running buffer containing dithiothreitol (2 mM) and sodium thioglycolate (1 mM). The resuspended pellets were then loaded onto the gel and electrophoresed at +70 V overnight (13 h).

$\alpha$ HL heptamers containing different numbers of mutant subunits were separated in the gel based on their electrophoretic mobilities which were determined by the number of octa-aspartate (D8) tails present (i.e., each cysteine-bearing mutant subunit contained a D8 tail). Hence, the top band corresponded to homoheptamers bearing no octa-aspartate tail (i.e., (AG)<sub>7</sub> or (MK)<sub>7</sub>), the second band corresponded to heteroheptamers bearing a single octa-aspartate tail (i.e., (mutant-D8)<sub>1</sub>(AG)<sub>6</sub> or (mutant-D8)<sub>1</sub>(MK)<sub>6</sub>) and so on, with consecutive bands having an aspartate-tail-free subunit replaced with a mutant subunit bearing an octa-aspartate tail. In this work, the second band from the top contained the desired protein pore with a single cysteine residue.

To extract the protein pores, the gel was dried under vacuum onto Whatman 3MM filter paper for 5 h at 50 °C. The dried gel was then exposed to photographic film (Kodak Bio Max MR autoradiography film) for 8 h and the developed film was used to locate the target protein bands in the gel. The desired protein bands were excised and

rehydrated in TE buffer (300  $\mu$ L; Tris-HCl (10 mM), DTT (0.5 mM), EDTA (1 mM), pH 8.0) for 1 h at room temperature. The paper was then removed, and the gel was crushed with a plastic pestle. The resulting suspension was filtered through a 0.2  $\mu$ m hydrophilic membrane filter (Proteus Mini Clarification Spin Column, Genron). The filtrate was stored in 10  $\mu$ L aliquots at -80 °C.

### Single-channel electrical recordings

Single-channel recordings were carried out in a planar bilayer apparatus as previously described<sup>38</sup>. A single  $\alpha$ HL pore was allowed to insert into the bilayer. Aldehyde substrates were introduced from the trans compartment. Experiments were conducted using recording buffer containing 2 M KCl, 200 mM PIPES and 20  $\mu$ M EDTA titrated to pH 6.8. Aldehyde solutions were prepared with recording buffer. Single-channel recordings were conducted with a coverslip placed atop the recording chamber.

Ionic currents were recorded by using a patch clamp amplifier (Axopatch 200B, Molecular Devices), and filtered with a low-pass Bessel filter (80 dB/decade) with a corner frequency of 10 kHz. Signals were digitised with a Digidata 1320 A digitizer (Molecular Devices) at an acquisition frequency of 50 kHz. The current traces were processed with Clampfit 10.7 (Molecular Devices). Current traces were idealised by using Clampfit 10.7 (Molecular Devices). The idealised data were analysed with QuB 2.0 software ([www.qub.buffalo.edu](http://www.qub.buffalo.edu))<sup>39</sup>. Dwell time analysis and rate constant determinations were performed by using the maximum interval likelihood algorithm of QuB<sup>40</sup>.

### *Arthrobacter cholorphenolicus* choline oxidase (AcCO6) preparation

The plasmid pET28a-AcCO6 encoding an engineered *Arthrobacter cholorphenolicus* choline oxidase (AcCO6) was a kind gift from Nicholas J. Turner<sup>26</sup>. The plasmid was transformed in *E. coli* BL21(DE3) competent cells for expression, which were cultured in LB media (10 mL) containing 50  $\mu$ g/mL kanamycin at 37 °C, 220 RPM overnight. The overnight culture was transferred to autoinduction media (1 L, LB based, Formedium, #AIMLB0110) containing 50  $\mu$ g/mL kanamycin and grown at 220 RPM for 16 hours at 20 °C. Cells were harvested by centrifugation at 4000  $\times$  g for 20 min at 4 °C. The pellet was resuspended in 50 mL pre-chilled Buffer A (100 mM KPi pH 7.8, 300 mM KCl, 20 mM imidazole) supplemented with 250 U nucleases (Pierce™ Universal Nuclease for Cell Lysis, #88700). Cells were lysed by sonication (10 s pulse, 10 s pause, 3 min duration) on ice, followed by centrifugation at 40,000  $\times$  g for 30 min at 4 °C. The supernatant was filtered through a 0.45  $\mu$ m syringe filter and applied to a 5 mL HisTrap™ FF column (Buffer A: 100 mM KPi pH 7.8, 300 mM KCl, 20 mM imidazole; Buffer B: 100 mM KPi pH 7.8, 300 mM KCl, 1000 mM imidazole). Fractions containing His-tagged AcCO6 were pooled and dialysed against a buffer containing 100 mM KPi pH 7.8 at 4 °C overnight. The His-tagged AcCO6 was concentrated, aliquoted, flash frozen with liquid N<sub>2</sub>, and stored at -80 °C.

### Differential sensing of aldehydes and alcohols

The control aldehyde-alcohol mixture was prepared as follows: propanal, butanal, 1-pentanol, 1-hexanol (5 mM each) were dissolved in oxygen-saturated KPi buffer (100 mM, pH 7.8). The treated mixture was prepared as follows: AcCO6 (0.1 mg/mL for each mM of alcohol) was added to the control aldehyde-alcohol mixture and shaken overnight at 300 RPM at 30 °C. After incubation, both the control and treated mixtures were adjusted to pH 6.8, and diluted 4-fold with buffer (2.66 M KCl, 200 mM PIPES at pH 6.8, 20  $\mu$ M EDTA) (i.e., final concentration of 2 M KCl).

Either the control (100  $\mu$ L) or treated mixture (100  $\mu$ L) was introduced into the trans chamber. As the control mixture was not treated with AcCO6, only propanal and butanal could be detected.

Single-channel recordings with the treated mixture revealed additional events with conductance levels corresponding to hemithioacetal adducts derived from pentanal and hexanal. This confirmed the oxidation of alcohols to aldehydes after AcCO<sub>6</sub> treatment.

**Note added in proof:** During the production of this paper, we reported complementary findings<sup>41</sup> on the diastereoselectivity of reactions between a nanoreactor cysteine and aromatic aldehydes. These results further confirm that protein nanopores can discriminate hemithioacetal diastereomers at the single-molecule level.

## Data availability

Data supporting the findings of this study are shown in the main text, Supplementary Information and Source data. Data needed to reproduce the machine learning results are available in the repository: <https://doi.org/10.5281/zenodo.16949339>. Source data are provided with this paper.

## Code availability

Codes needed to reproduce the machine learning results are available in the repository: <https://doi.org/10.5281/zenodo.16949339>.

## References

1. Sutaria, S. R. et al. Lipid peroxidation produces a diverse mixture of saturated and unsaturated aldehydes in exhaled breath that can serve as biomarkers of lung cancer—a review. *Metabolites* **12**, 561 (2022).
2. Schauer, J. J., Kleeman, M. J., Cass, G. R. & Simoneit, B. R. T. Measurement of emissions from air pollution sources. 3. C<sub>1</sub>-C<sub>29</sub> organic compounds from fireplace combustion of wood. *Environ. Sci. Technol.* **35**, 1716–1728 (2001).
3. Moreira, M. T. G., Pereira, P. R., Aquino, A., Conte-Junior, C. A. & Paschoalin, V. M. F. Aldehyde accumulation in aged alcoholic beer: addressing acetaldehyde impacts on upper aerodigestive tract cancer risks. *Int. J. Mol. Sci.* **23**, 14147 (2022).
4. Smolinska, A. et al. Volatile metabolites in breath strongly correlate with gut microbiome in CD patients. *Anal. Chim. Acta* **1025**, 1–11 (2018).
5. Rudnicka, J., Kowalkowski, T. & Buszewski, B. Searching for selected VOCs in human breath samples as potential markers of lung cancer. *Lung Cancer* **135**, 123–129 (2019).
6. Hicks, L. C. et al. Analysis of exhaled breath volatile organic compounds in inflammatory Bowel disease: a pilot study. *J. Crohns Colitis* **9**, 731–737 (2015).
7. Ibrahim, W. et al. Visualization of exhaled breath metabolites reveals distinct diagnostic signatures for acute cardiorespiratory breathlessness. *Sci. Transl. Med.* **14**, eab15849 (2022).
8. Ibrahim, W. et al. Diagnosis of COVID-19 by exhaled breath analysis using gas chromatography–mass spectrometry. *ERJ Open Res.* **7**, 00139-2021 (2021).
9. Su, R. et al. Mass spectrometry for breath analysis. *Trends Anal. Chem.* **158**, 116823 (2023).
10. Drabińska, N. et al. A literature survey of all volatiles from healthy human breath and bodily fluids: the human volatilome. *J. Breath. Res.* **15**, 034001 (2021).
11. Fuchs, P., Loeseken, C., Schubert, J. K. & Miekisch, W. Breath gas aldehydes as biomarkers of lung cancer. *Int. J. Cancer* **126**, 2663–2670 (2010).
12. Guadagni, R. et al. Solid-phase microextraction–gas chromatography–mass spectrometry method validation for the determination of endogenous substances: urinary hexanal and heptanal as lung tumor biomarkers. *Anal. Chim. Acta* **701**, 29–36 (2011).
13. Bayley, H., Luchian, T., Shin, S.-H. & Steffensen, M. B. Single-molecule covalent chemistry in a protein nanoreactor. In *Single Molecules and Nanotechnology* (eds Rigler, R. & Vogel, H.) (Springer, 2008).
14. Wu, H.-C. & Bayley, H. Single-molecule detection of nitrogen mustards by covalent reaction within a protein nanopore. *J. Am. Chem. Soc.* **130**, 6813–6819 (2008).
15. Boersma, A. J. & Bayley, H. Continuous stochastic detection of amino acid enantiomers with a protein nanopore. *Angew. Chem. Int. Ed.* **51**, 9606–9609 (2012).
16. Boersma, A. J., Brain, K. L. & Bayley, H. Real-time stochastic detection of multiple neurotransmitters with a protein nanopore. *ACS Nano* **6**, 5304–5308 (2012).
17. Ramsay, W. J. & Bayley, H. Single-molecule determination of the isomers of D-glucose and D-fructose that bind to boronic acids. *Angew. Chem. Int. Ed.* **57**, 2841–2845 (2018).
18. Liu, Y. et al. Nanopore identification of alditol epimers and their application in rapid analysis of alditol-containing drinks and healthcare products. *J. Am. Chem. Soc.* **144**, 13717–13728 (2022).
19. Fan, P. et al. Nanopore analysis of Cis-diols in fruits. *Nat. Commun.* **15**, 1969 (2024).
20. Wang, K. et al. Unambiguous discrimination of all 20 proteinogenic amino acids and their modifications by nanopore. *Nat. Methods* **21**, 92–101 (2024).
21. Zhang, M. et al. Real-time detection of 20 amino acids and discrimination of pathologically relevant peptides with functionalized nanopore. *Nat. Methods* **21**, 609–618 (2024).
22. Choi, L. S., Mach, T. & Bayley, H. Rates and stoichiometries of metal ion probes of cysteine residues within ion channels. *Biophys. J.* **105**, 356–364 (2013).
23. Hilal, S. H., Bornander, L. L. & Carreira, L. A. Hydration equilibrium constants of aldehydes, ketones and quinazolines. *QSAR Comb. Sci.* **24**, 631–638 (2005).
24. Lienhard, G. E. & Jencks, W. P. Thiol addition to the carbonyl group. Equilibria and kinetics. *J. Am. Chem. Soc.* **88**, 3982–3995 (1966).
25. Daw, G., Regan, A. C., Watt, C. I. F. & Wood, E. Steric effects and mechanism in the formation of hemiacetals from aliphatic aldehydes. *J. Phys. Org. Chem.* **26**, 1048–1057 (2013).
26. Heath, R. S. et al. An engineered alcohol oxidase for the oxidation of primary alcohols. *ChemBioChem* **20**, 276–281 (2019).
27. El-Maghrabey, M. H. et al. Comprehensive greenness evaluation of the reported chromatographic methods for aldehydes determination as clinical biomarkers and food quality indicators. *Trends Anal. Chem.* **171**, 117548 (2024).
28. Poli, D. et al. Determination of aldehydes in exhaled breath of patients with lung cancer by means of on-fiber-derivatization SPME–GC/MS. *J. Chromatogr. B* **878**, 2643–2651 (2010).
29. Hamidi, S. Recent advances in solid-phase extraction as a platform for sample preparation in biomarker assay. *Crit. Rev. Anal. Chem.* **53**, 199–210 (2023).
30. Pleil, J. D., Wallace, M. A. G., Davis, M. D. & Matty, C. M. The physics of human breathing: flow, timing, volume, and pressure parameters for normal, on-demand, and ventilator respiration. *J. Breath. Res.* **15**, 042002 (2021).
31. Czekalska, M. A. et al. A droplet microfluidic system for sequential generation of lipid bilayers and transmembrane electrical recordings. *Lab Chip* **15**, 541–548 (2015).
32. Yu, L.-Q. et al. A gate-opening controlled metal-organic framework for selective solid-phase microextraction of aldehydes from exhaled breath of lung cancer patients. *Microchim. Acta* **185**, 307 (2018).
33. Noreña-Caro, D. & Álvarez-Láinez, M. Functionalization of polyacrylonitrile nanofibers with β-cyclodextrin for the capture of formaldehyde. *Mater. Des.* **95**, 632–640 (2016).
34. Bo, Z., Lim, Z. H., Duarte, F., Bayley, H. & Qing, Y. Mobile molecules: reactivity profiling guides faster movement on a cysteine track. *Angew. Chem. Int. Ed.* **62**, e2023008 (2023).

35. DREAMER, D., AKESON, M. & BRANTON, D. Three decades of nanopore sequencing. *Nat. Biotechnol.* **34**, 518–524 (2016).
36. Schober, L. et al. Enzymatic reactions towards aldehydes: an overview. *Flavour Fragr. J.* **38**, 221–242 (2023).
37. Braha, O. et al. Designed protein pores as components for biosensors. *Chem. Biol.* **4**, 497–505 (1997).
38. Montal, M. & Mueller, P. Formation of biomolecular membranes from lipid monolayers and a study of their electrical properties. *Proc. Natl. Acad. Sci. USA* **69**, 3561–3566 (1972).
39. Nicolai, C. & Sachs, F. Solving ion channel kinetics with the QuB software. *Biophys. Rev. Lett.* **08**, 191–211 (2013).
40. Qin, F., Auerbach, A. & Sachs, F. Estimating single-channel kinetic parameters from idealized patch-clamp data containing missed events. *Biophys. J.* **70**, 264–280 (1996).
41. Lim, Z. H., Bo, Z., Armstrong, E., Bayley, H. & Qing, Y. Unexpected diastereoselective chemistry on a 2D protein surface. *Chem.* 102717, <https://doi.org/10.1016/j.chempr.2025.102717> (2025).

## Acknowledgements

This research was supported by the Bill & Melinda Gates Foundation, a European Research Council Advanced Grant (SYNTISU), and a European Research Council Starting Grant (NANOPRO) under the UK Research and Innovation (UKRI) Guarantee (EP/Z000351/1). Y.Y. is supported by an EPSRC Postdoctoral Fellowship (UKRI939). G.W. thanks the support of UKRI for the Horizon Europe Guarantee funding of a Marie Skłodowska-Curie Actions Postdoctoral Fellowship (EP/Z001366/1). We thank Prof. Nicholas J. Turner from the University of Manchester for providing the AcCO6-encoding plasmid. We thank Steven Barry at the Physical and Theoretical Chemistry Laboratory, University of Oxford, for fabricating the recording chambers.

## Author contributions

H.B. and Y.Q. conceived and supervised the project. L.E.M. and Z.H.L. prepared the proteins and conducted the single-channel recording experiments. Z.B. prepared the script for machine learning and performed the molecular dynamics simulations. Y.Y. prepared the AcCO6 enzyme. G.W. carried out parallel sensing experiments. L.E.M., Z.H.L., Z.B. and Y.Q. performed the data analysis. L.E.M., Z.H.L., Y.Y., Z.B., H.B. and Y.Q. wrote the manuscript.

## Competing interests

H.B. is the founder of, a consultant for and a shareholder of Oxford Nanopore Technologies, a company engaged in the development of nanopore sensing and sequencing technologies. L.E.M., Z.H.L., H.B. and Y.Q. have filed patents describing the engineered nanopores and their applications in small-molecule covalent sensing. Z.B. and Y.Y. are listed as contributors to the IP. The remaining authors declare no competing interests.

## Additional information

**Supplementary information** The online version contains supplementary material available at <https://doi.org/10.1038/s41467-025-64184-6>.

**Correspondence** and requests for materials should be addressed to Hagan Bayley or Yujia Qing.

**Peer review information** *Nature Communications* thanks the anonymous reviewers for their contribution to the peer review of this work. A peer review file is available.

**Reprints and permissions information** is available at <http://www.nature.com/reprints>

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2025