

Cite this: *Chem. Sci.*, 2021, 12, 9318

All publication charges for this article have been paid for by the Royal Society of Chemistry

## Aromatic side-chain flips orchestrate the conformational sampling of functional loops in human histone deacetylase 8†

Vaibhav Kumar Shukla,<sup>‡a</sup> Lucas Siemons,<sup>‡a</sup> Francesco L. Gervasio<sup>bcd</sup> and D. Flemming Hansen<sup>‡\*a</sup>

Human histone deacetylase 8 (HDAC8) is a key hydrolase in gene regulation and an important drug-target. High-resolution structures of HDAC8 in complex with substrates or inhibitors are available, which have provided insights into the bound state of HDAC8 and its function. Here, using long all-atom unbiased molecular dynamics simulations and Markov state modelling, we show a strong correlation between the conformation of aromatic side chains near the active site and opening and closing of the surrounding functional loops of HDAC8. We also investigated two mutants known to allosterically downregulate the enzymatic activity of HDAC8. Based on experimental data, we hypothesise that I19S-HDAC8 is unable to release the product, whereas both product release and substrate binding are impaired in the S39E-HDAC8 mutant. The presented results deliver detailed insights into the functional dynamics of HDAC8 and provide a mechanism for the substantial downregulation caused by allosteric mutations, including a disease causing one.

Received 6th April 2021  
Accepted 27th May 2021

DOI: 10.1039/d1sc01929e

rsc.li/chemical-science

### Introduction

Acetylation of lysine side chains occurs as a co-translation or post-translational modification of proteins and was first identified in histone proteins as a dynamic and reversible process.<sup>1</sup> The acetylation and de-acetylation reactions are catalysed by histone acetyl-transferases (HATs) and histone deacetylase (HDAC) enzymes, respectively.<sup>1</sup> Acetylation of histone lysine side chains is often associated with transcriptional activation, whereas de-acetylation often leads to closed chromatin structures and repression of transcription of the underlying genes.<sup>2</sup> It is therefore not surprising that aberrant acetylation or deacetylation patterns have been linked to several diseases including leukaemia, lymphomas, and neurodegenerative disorders.<sup>3–8</sup> Substantial efforts are now made to target HDACs for the treatment of such diseases with some HDAC inhibitors already approved by the FDA and several in preclinical trials.

HDACs are traditionally divided into four classes based on sequence similarities. Class I (HDAC-1, -2, -3, and HDAC8), class

II (HDAC-4, -5, -6, -7, -9, and HDAC10), and class III (SIRT1-7) have sequence similarity to yeast Rpd3, Hda1, and Sir2, respectively, whereas class IV (HDAC11) shares sequence similarity with both class I and II proteins.<sup>2</sup> In cancers, class I HDACs act as major epigenetic players and are linked to deregulated expression or interactions with transcription factors critical to tumorigenesis.<sup>9</sup> Histone deacetylase 8, a class I HDAC, is also implicated in other diseases, including X-linked intellectual disability and parasitic infections.<sup>3,10,11</sup> Additionally, genetic mutations leading to impaired HDAC8 activity are reported in patients with Cornelia de Lange syndrome (CdLS).<sup>7,8,12–16</sup>

HDAC8 is a unique class I HDAC in terms of its structure, activity, and regulation. Unlike the other three members of class I, HDAC8 is constitutively active, whereas HDAC1 and HDAC2 show activity that increases dramatically once present in multi-protein complexes, and HDAC3 is inactive in isolation.<sup>2</sup> Moreover, HDAC1, HDAC2, and HDAC3 have serine phosphorylation sites in their flexible C-terminal regulatory tail, which upon phosphorylation activates the enzyme. In contrast, HDAC8 has a regulatory serine phosphorylation site at position 39, in the catalytic domain, and its phosphorylation inactivates the enzyme by stabilising an inactive state.<sup>2,17</sup> Recently, in HDAC1, HDAC2, and HDAC3, the region corresponding to the S39 site in HDAC8 was shown to act as a binding platform in holoenzyme HDAC complexes, suggesting a general mechanism for regulation across class I HDACs.<sup>17–20</sup>

The overall fold of HDAC8, as observed in substrate- and inhibitor-HDAC8 complexes, is comprised of an eight stranded

<sup>a</sup>Department of Structural and Molecular Biology, Division of Biosciences, University College London, London, WC1E 6BT, UK. E-mail: d.hansen@ucl.ac.uk

<sup>b</sup>Department of Chemistry, University College London, London, WC1E 6BT, UK

<sup>c</sup>Pharmaceutical Sciences, University of Geneva, Geneva CH-1211, Switzerland

<sup>d</sup>Institute of Pharmaceutical Sciences of Western Switzerland, University of Geneva, Geneva CH-1211, Switzerland

† Electronic supplementary information (ESI) available: Supporting Fig. S1–S11, supporting Movies S1 and S2. See DOI: 10.1039/d1sc01929e

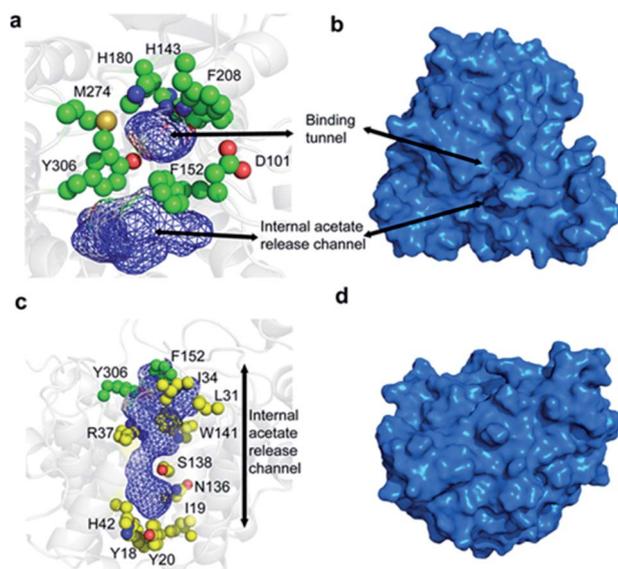
‡ These authors contributed equally to this work.



$\beta$ -sheet, which is surrounded by 11 helices and two helical turns.<sup>21–26</sup> The active site is formed by seven loops, L1 to L7, connecting these secondary structure elements. The substrate-binding tunnel is formed by residues D101, H142, H143, G151, F152, H180, F208, M274, and Y306 with the active-site  $\text{Zn}^{2+}$  located at the base of the tunnel, Fig. 1a and b. These residues are conserved across the class I HDACs, except for M274, which is a leucine in the other class I family members. For class II members HDAC6 and HDAC10 a lysine and a glutamate is present, respectively, whereas a leucine is present in other class II HDAC family members at this position.<sup>21</sup>

It has been proposed that, in the unbound form, the active-site  $\text{Zn}^{2+}$  adopts a 5-coordinate square pyramidal geometry with D178, H180, and D267 and two water molecules.<sup>27</sup> Upon substrate or inhibitor binding, one of the water molecules is displaced by the side-chain carbonyl of the substrate/inhibitor. The  $\text{Zn}^{2+}$  ion and H142 activate a water molecule for a nucleophilic attack on the carboxyl of the acetylated lysine side chain and the reaction intermediate has been proposed to be stabilised by Y306, H142, H143, and  $\text{Zn}^{2+}$ . Subsequently, the hydrolysis is completed by H143 that protonates the lysine side chain.<sup>27</sup>

Following hydrolysis, the acetate product has been proposed to be released *via* an internal channel adjacent to the substrate-binding tunnel.<sup>21,28</sup> This ‘release channel’ is lined by residues I34, F152, Y306 from the top and Y18, I19, Y20, H42, N136 from the bottom, Fig. 1c and d, with Y306, W141 and F152 located at



**Fig. 1** Structure of HDAC8 highlighting the binding tunnel and the proposed acetate release channel. (a) A top view of the binding tunnel and internal acetate release channel. Residues forming the wall of the binding tunnel and those present at the junction of the substrate-binding tunnel and the acetate release channel are shown in green. (b) Exterior surface representation depicting the substrate-binding tunnel in the same orientation as shown in (a). (c) Side view of the internal acetate release channel. Residues surrounding only the acetate release channel are shown in yellow. (d) Exterior surface representation of the orientation in (c).

the intersection between the substrate-binding and acetate release channel. The internal acetate release channel has been hypothesised to divide into sub-channels with R37 and S138 present at the junction and R37 playing an important role in stabilising the negatively charged acetate for effective release. Despite the availability of many crystal structures of HDAC8 in complex with different substrates and inhibitors,<sup>21–26</sup> a high-resolution structure of the unbound form of HDAC8, or any other class I HDACs, is not yet available. Also, many aspects of HDAC8 regulation and function still remain elusive,<sup>29–32</sup> in particular the mechanism of product release as well as the role of the dynamic loops, L1–L7, for function and how these are connected to the active site.

In this work we performed 10  $\mu\text{s}$ -long unbiased molecular dynamics simulations of wild-type HDAC8 and two mutants I19S and S39E, which are known to allosterically downregulate activity. We observe that dynamics about the  $\chi_1$  dihedral angle of aromatic residues near the active site, Y306, H143, and F152, orchestrate the conformational sampling of the functional loops of HDAC8. In both mutants, I19S and S39E, the conformational sampling of the functional loops is significantly affected. In I19S-HDAC8 the dynamics of Y306 and the L6 loop is completely inhibited and a  $\sim 20$  Å allosteric path from I19 to Y306 is identified, which passes through several sites that have been identified to be essential for activity. The S39E mutant leads to a substantial stabilisation of all the aromatic side chains near the active site, which effectively traps each of the functional loops in one conformation. Overall, the results show how the active site of HDAC8 is connected to the conformational sampling of functional loops and also how mutations affect the sampled conformations and activity.

## Methods

### Experimental design

Long (10  $\mu\text{s}$ ) all-atom unbiased molecular dynamics simulations were used to characterise allosteric communications within the free form of the HDAC8 enzyme. The wild type HDAC8 as well as two mutants, I19S-HDAC8 and S39E-HDAC8, known to allosterically downregulate enzymatic activity were investigated to gain insight into the mechanism of down-regulation. Markov state modelling was used to provide insight into the correlation between relevant motions observed.

### Structure preparation and simulation setup

To prepare the wild-type HDAC8 model the crystal structure (PDB: 1T64)<sup>21</sup> was used, with the missing loops modelled in using MODELLER<sup>33</sup> and the trichostatin A molecules removed. Following this, hydrogens were added and the protonation states were determined using MolProbity.<sup>34</sup> The system was parameterised using the amber99SB\*-ILDN forcefield with TIP3P water.<sup>35,36</sup> To maintain the  $\text{Zn}^{2+}$  and  $\text{K}^+$  ions bound during the simulation, a harmonic potential centred around 2.25 Å, with a force constant of 3.960 J mol<sup>-1</sup>, was placed between the  $\text{Zn}^{2+}$  cation and H180 N <sup>$\delta$ 1</sup> and between the potassium distal to the active site and the carbonyl groups of V195 and T192. An *in*



*vacuo* energy minimisation was carried out.<sup>37</sup> Subsequently the system was solvated in a dodecahedral box with a volume of 458.83 nm<sup>3</sup> and neutralised by adding K<sup>+</sup> counter-ions followed by an additional 19 K<sup>+</sup> and 19 Cl<sup>-</sup> ions (70 mM). Following this, a second energy minimisation was carried out and two density equilibrations were performed. The first used a Berendsen barostat and thermostat for 100 ps.<sup>37</sup> The second equilibration was carried out with the Parrinello–Rahman barostat<sup>38</sup> with a Nosé–Hoover thermostat<sup>39,40</sup> for 100 ps. Finally, a 2 ns NVT step was carried out to equilibrate the systems using the final run conditions with coupling to a velocity-rescaling thermostat.<sup>41</sup> During all steps, the bond lengths were constrained using the LINCS algorithm<sup>42</sup> and electrostatics were calculated using the PME algorithm with a 1.2 nm radius.<sup>43</sup> The van der Waals interactions were determined using a standard 1.2 nm cut-off scheme. The length of each time step was 2 fs and all simulations were carried out at a temperature of 300 K. The simulations were carried out using Gromacs.<sup>44</sup>

The I19S and S39E simulations were prepared similar to the wild-type simulations after introducing the mutation. In order to obtain a simulation for the reverse mutation, I19Sr, the final frame from the I19S simulation was taken and S19 was mutated back to isoleucine. The waters were removed and the model was prepared as described above. Finally, the wild-type HDAC8, S39E-HDAC8 simulations were run for 10 μs, the I19S-HDAC8 simulation for 8 μs, whereas the I19Sr-HDAC8 simulation was run for 2.5 μs. The lengths of the simulations were tailored to the characteristic period of the observed conformational changes and the convergence of the sampling was also monitored through Markov state analysis. An equilibrium period of 2 μs, 1.8 μs, and 2.5 μs was used for the wild-type, S39E-HDAC8, and I19S-HDAC8 simulation, respectively. This equilibrium period was not used for any of the analyses presented and the length of the equilibrium period was determined based on the RMSD to the start structure, Fig. S1.† The loop, between K202 to V217 was not considered in the presented analysis, because convergence of this loop could not be fully justified.

### Markov state analysis

For the wild-type HDAC8 simulation, the correlation between the  $\chi_1$  dynamics of H143 and Y306 was characterised using a Markov state model. Initially the  $\chi_1$  angle of H143 and Y306 was calculated for every 100 ps throughout the simulation and subsequently one of the six states, {H143(g-)Y306(g-), H143(g-)Y306(t), H143(g-)Y306(g+), H143(t)Y306(g-), H143(t)Y306(t), H143(t)Y306(g+)} was assigned. Thereafter, the transition matrix, **T**, was calculated as described previously.<sup>45</sup> The convergence and the appropriate lag time were assessed using the Chapman–Kolmogorov equation. First the six eigenvalues,  $\lambda_i$ , of the transition matrix were calculated as a function of lag time,  $\tau_{\text{lag}} = n \times 100$  ps and subsequently the slowest relaxation time,  $t_i = -\tau_{\text{lag}}/\log(\lambda_i)$ , was calculated for different lag times, Fig. S2.† According to the Chapman–Kolmogorov equation the Markov model was assumed converged at the lag time, where the slowest relaxation time becomes independent of lag time,

which was 2.8 ns. In order to calculate the transition rates the symmetrised transition matrix was used,  $\tilde{\mathbf{T}}$ , and only transitions where either the side chain of H143 or Y306 flips are considered. Specifically, a least-squares minimisation of  $\|\tilde{\mathbf{T}} - \exp(-\tau_{\text{lag}}\mathbf{\Gamma})\|_2^2$ , with respect to the rate constants was performed, where  $\mathbf{\Gamma}$  is the rate matrix holding the rates.

### Calculation of Hellinger divergences

The residue-specific distributions of  $\{\varphi, \psi\}$  and  $\chi_1$  in the wild-type and I19S simulation were compared using a Hellinger divergence measure. Specifically, probability distributions for  $\{\varphi, \psi\}$  and  $\chi_1$ ,  $p(\varphi, \psi)$  and  $p(\chi_1)$  were first estimated using a von Mises distribution in Fourier space with  $\kappa = 120$  and 120 bins over the domain  $\varphi, \psi, \chi_1 \in [-180^\circ, 180^\circ]$ . Subsequently the Hellinger divergence distances,  $H$ , was calculated between the  $\{\varphi, \psi\}$  distribution of the wild-type and the I19S simulations, and between the  $\chi_1$  distribution for the wild-type and I19S simulations by

$$1 - H_{\Omega}^2 = \int d\Omega \sqrt{p_{\text{WT}}(\Omega)p_{\text{I19S}}(\Omega)}$$

with  $\Omega$  being either  $\{\varphi, \psi\}$  or  $\chi_1$ . Finally, the total Hellinger distance,  $H_{\text{T}}$ , was calculated as

$$H_{\text{T}}^2 = H_{\{\varphi, \psi\}}^2 + H_{\chi_1}^2$$

A Hellinger divergence was only considered significant if it was larger than the average divergence obtained for a block-analysis, {{0 μs, 2.5 μs}, {2.5 μs, 5.0 μs}, {5.0 μs, 7.5 μs}, {7.5 μs, 10.0 μs}}, of the wild-type simulation.

All analyses was carried out using MDAnalysis,<sup>46</sup> and Python 3.6 or Python 3.8 with the Numpy and Matplotlib libraries.

## Results

### Y306 and H143 each show two stable side-chain conformations

A 10 μs-long atomistic molecular dynamics (MD) simulation of wild-type HDAC8 reveals reversible conformational changes of residues involved in substrate-binding and catalysis with an average lifetime of ~2 micro-second. For Y306, which is crucial for both substrate binding and catalysis, two stable states are sampled, Fig. 2a.

The major *gauche-* state ( $\chi_1 \approx 300^\circ$ ; g-) is characterised by a ~6.2 Å distance between Y306-O<sup>c</sup> and the active Zn<sup>2+</sup> ion, whereas the minor *gauche+* state ( $\chi_1 \approx 60^\circ$ ; g+) is characterised by a substantial longer Y306-O<sup>c</sup> to Zn<sup>2+</sup> distance of ~11.5 Å and the tyrosine side chain pointing towards the L6 loop, Fig. 2b. The flips between the *gauche+* and *gauche-* states of Y306 are strongly correlated with a 5-to-7 Å displacement of the L6 loop, wherein M274 is located, Fig. 2c. Loop L6 and M274 form a part of the substrate-binding tunnel and in the Y306(g+) state the tunnel is substantially wider and solvent exposed, thus showing that the *gauche-* to *gauche+* flip of Y306 acts as a switch to open the substrate-binding tunnel, Fig. 3. Further examination of the Y306(g+) state shows that it is closely associated with



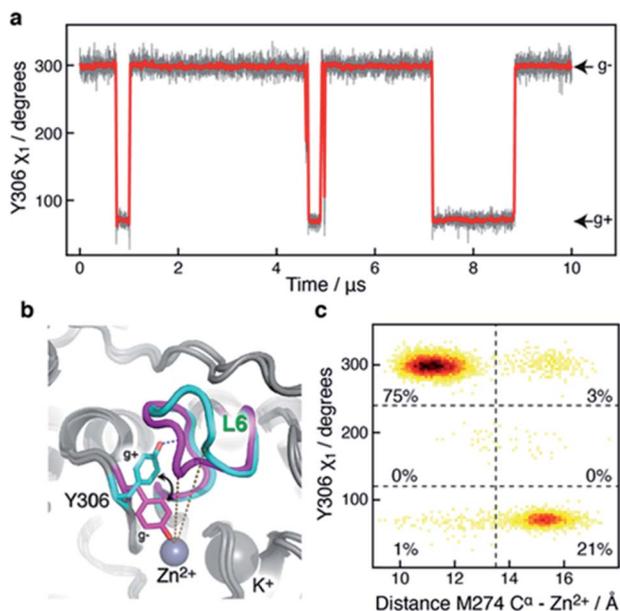


Fig. 2 Motion of the active-site Y306 residue correlates with L6 conformation. (a) The Y306  $\chi_1$  angle as a function of simulation time (grey) and shown every nanosecond. The red line is a moving average over a window of 20 ns. Two stable and distinct conformations are observed, Y306(g<sup>-</sup>) and Y306(g<sup>+</sup>). (b) The Y306(g<sup>-</sup>) and Y306(g<sup>+</sup>) conformations are shown in magenta and cyan, respectively. The g<sup>+</sup> conformation is stabilised by a hydrogen-bond between Y306-OH and I269 CO (blue dotted line). Moreover, the Y306  $\chi_1$  flip is associated with a large change in the L6 loop conformation. The active Zn<sup>2+</sup> and the regulatory K<sup>+</sup> ions are shown as grey spheres. The distance between Zn<sup>2+</sup> and M274 C $\alpha$  in two stable conformations is shown as dotted lines. (c) Two-dimensional histogram showing that the distance between M274 C $\alpha$  within the L6 loop and Zn<sup>2+</sup> correlates strongly with the  $\chi_1$  dihedral angle of Y306.

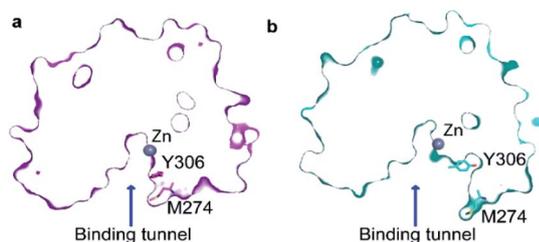


Fig. 3 Transition from Y306(g<sup>-</sup>) to Y306(g<sup>+</sup>) leads to an open state of the active site. (a) The cut-away surface view of a representative structure of the substrate-binding tunnel in the Y306(g<sup>-</sup>) state. (b) The cut-away surface view of a representative structure of the substrate-binding tunnel and active site in the Y306(g<sup>+</sup>) state, where the active site and substrate-binding tunnel are exposed to the bulk solvent.

a hydrogen bond between Y306-OH and I269-CO, Fig. 2b, which implies that this hydrogen bond may stabilise the Y306(g<sup>+</sup>) state. Moreover, in the Y306(g<sup>+</sup>) state a salt-bridge between D233 and R353 is fully formed, whereas in the Y306(g<sup>-</sup>) state this salt-bridge is only partly formed, Fig. S3.† The D233–R353 salt bridge, in turn, stabilises the L6 loop in the Y306(g<sup>+</sup>) state *via* additional charge–charge interactions between R353 and

Cys275-CO as well as between R353 and Pro273-CO. In addition to the Y306(g<sup>-</sup>) and Y306(g<sup>+</sup>) states, there is also a very low-populated and short-lived state with the Y306  $\chi_1$  in a *trans* state (~0.5% population), where the Y306 side chain points towards the L1 loop.

Of particular interest is that the substrate-binding tunnel is substantially widened and the active site exposed to the bulk solvent when the L6 loop is distant to the Zn<sup>2+</sup>, Fig. 3 and S4.† Such a state, with the exposed active site and the widened substrate-binding tunnel, will likely be able to rapidly release products. A movie for visually comparing the structure in the “closed” and “open” states, and with a transition from surface to cut-away, is shown in ESI Movie S1.†

Evidence of conformational flexibility around Y306 was previously observed in a 10 ns simulation of Y306F-HDAC8 (ref. 47) and the *gauche*<sup>+</sup> and *trans* conformations have been observed in crystal structures of related enzymes, thereby providing evidence for the states of Y306 observed in the unbiased simulations of HDAC8 presented here. Specifically, in the crystal structure of acetylpolymine amidohydrolase (APAH; PDB 3Q9F), *gauche*<sup>-</sup> and *trans* were observed for the corresponding residue, while in H976Y-HDAC4 the *gauche*<sup>-</sup> (PDB: 2VQV) and the *gauche*<sup>+</sup> (PDB: 2VQO) were observed.<sup>48,49</sup>

The active-site residue H143 is involved in the hydrolysis by forming a hydrogen-bond with the acetylated lysine N<sup>ε</sup> and thus stabilising the reaction intermediate. As observed for Y306, H143 also exists in a dynamic equilibrium between two

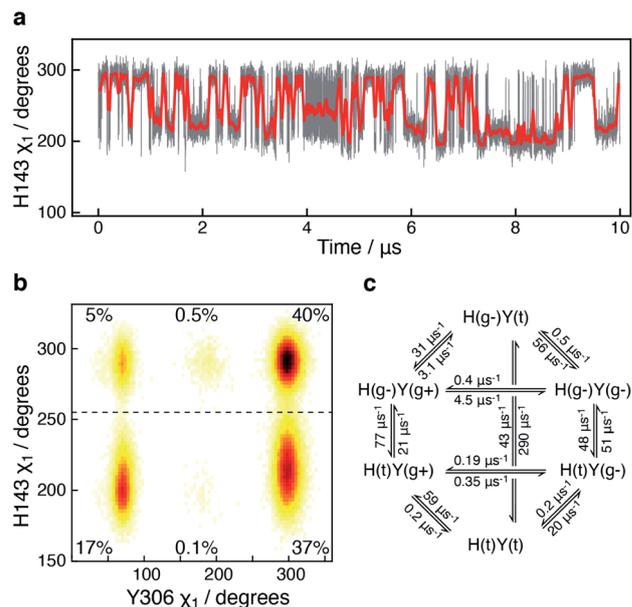


Fig. 4 Side-chain motion of H143 correlates with Y306 sampling: (a) the H143  $\chi_1$  angle as a function of simulation time (grey) and sampled every nanosecond; the red line is a moving average over a window of 20 ns. The H143 side chain samples the *gauche*<sup>-</sup> ( $\chi_1 \approx 300^\circ$ ; g<sup>-</sup>) and the *trans* ( $\chi_1 \approx 205^\circ$ ; t) states each with an average life-time of 11 ns. (b) A change in the Y306  $\chi_1$  angle alters the sampling of H143  $\chi_1$  such that H143(t) is stabilised when Y306  $\chi_1$  is in a *gauche*<sup>+</sup> conformation. (c) Markov state model derived for the H143 (H) and Y306  $\chi_1$  (Y) combined dynamics.



conformations that are characterised by different  $\chi_1$  angles, H143(g<sup>-</sup>) with  $\chi_1 \approx 300^\circ$  and H143(t) with  $\chi_1 \approx 205^\circ$ , Fig. 4.

In all crystal structures of HDAC8, bound to a substrate or inhibitor, only the H143(g<sup>-</sup>) is present.<sup>21–26</sup> When the H143 side chain flips from *gauche*<sup>-</sup> and *trans*, the entire L2 loop region (residues 86–103), as well as the L3 loop located between L2 and the active site (residues 142–154), move approximately 1 Å closer to the active site Zn<sup>2+</sup>, Fig. S5.† The functional L2 loop region contains D101, which is known to stabilise substrate binding.<sup>50,51</sup> Moreover, a correlation is observed between the sampling of the  $\chi_1$  of Y306 and H143, Fig. 4b. Specifically, when the Y306(g<sup>+</sup>) state is formed, H143  $\chi_1$  is stabilised in the *trans* conformation.

To gain further insight into how the state of Y306 influence the dynamics of H143, and *vice versa*, a Markov state model was generated from the 10 μs simulation, Fig. 4c (see Methods). The generated Markov model highlights the stabilisation of H143(t) when Y306  $\chi_1$  is in a *gauche*<sup>+</sup> conformation, a stabilisation that appears to be mainly caused by a slower rate from H143(t) to H143(g<sup>-</sup>) and a slightly faster rate from H143(g<sup>-</sup>) to H143(t). Also, the dynamics of Y306 is essentially frozen when H143  $\chi_1$  is in the *trans* state. Taken together these observations suggest that the motions of the two functional residues, Y306 and H143, greatly influence each other and the motion of each of these aromatic side chains correlate with the conformations of the functional loops.

### The conformation of F152 steers the L1 loop

In addition to the motions of Y306 and H143 involved in binding and catalysis, similar  $\chi_1$  motions were observed for F152, which forms a part of both the substrate-binding tunnel and the acetate release channel. F152  $\chi_1$  samples predominantly two states, F152(g<sup>-</sup>) and F152(t), with associated average  $\chi_1$  angles of *ca.* 300° and *ca.* 200°, respectively.

The  $\chi_1$  motion of F152 is substantially faster than those of H143 and Y306, with F152(g<sup>-</sup>) and F152(t) having an average life-time of 1.4 ns and 1.1 ns, respectively, during the 10 μs simulation, Fig. 5. Thus, there is a slight preference for the F152(g<sup>-</sup>) state (58%) over the F152(t) state (42%). Moreover, the F152(t) state is stabilised twice, for approximately 1 μs each time, Fig. 5a. The F152(g<sup>+</sup>) state is only present at 0.3%.

The F152 residue is found in the *gauche*<sup>-</sup> conformation in all human HDAC8 crystal structures solved so far, however, in the HDAC8 homolog from *Schistosoma mansoni* the corresponding phenylalanine residue, F151, is observed in a *trans* state,<sup>52</sup> see Fig. S6,† thereby providing evidence that the substantial amount of *trans* conformation observed in the unbiased simulations is realistic for HDAC8 without a substrate or an inhibitor bound.

The correlation between the conformation of the L1 loop and the  $\chi_1$  angle of F152 is striking, Fig. 5b. When F152 is in the *trans* state, the L1 loop moves approximately 3 Å away from the substrate-binding tunnel leading to an open state. In line with this movement is that the salt-bridge between K33 in L1 and D87 in L2, previously related to function,<sup>50</sup> predominantly forms in the F152(g<sup>-</sup>) state. These findings agree with the

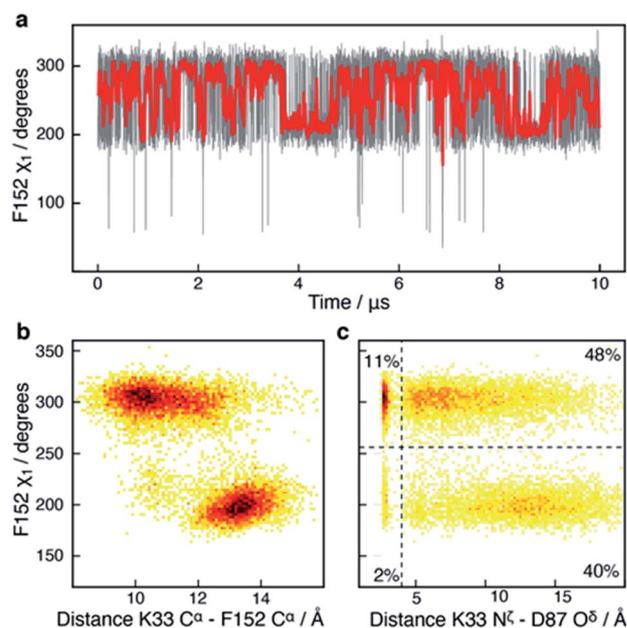


Fig. 5 Side-chain motion of F152 correlates with L1 sampling: (a) the F152  $\chi_1$  angle as a function of simulation time (grey) and sampled every nanosecond; the red line is a moving average over a window of 20 ns. The F152 side chain samples the *gauche*<sup>-</sup> ( $\chi_1 \approx 300^\circ$ ; g<sup>-</sup>) and the *trans* ( $\chi_1 \approx 200^\circ$ ; t) states each with an average life-time of about 1 ns. (b) Two-dimensional histogram showing that the  $\chi_1$  dihedral angle of F152 correlates strongly with the distance between K33 C $\alpha$  within the L1 loop and F152 C $\alpha$ ; a distance which has previously been used as a measure of the L1 conformation.<sup>37</sup> (c) Two-dimensional histogram showing that the distance between the side chains of K33 in L1 and D87 in L2 as well as ability to form a salt-bridge between the two side chains, correlate with the  $\chi_1$  of F152. In the F152(t) state, a salt-bridge between L1 and L2 is unlikely (5%), whereas in the F152(g<sup>+</sup>) state there is a 22% chance of a salt-bridge being formed.

crystal structure of the HDAC8 homolog from *Schistosoma mansoni*, where the corresponding phenylalanine is in *trans*, and the L1 loop is *ca.* 1 Å further away from the active site compared to human HDAC8.

Despite the motion about the F152  $\chi_1$  angle being fast, the conformation of other functional sites is correlated with the state of the F152 side chain. One example is R37, Fig. S7,† which has previously been suggested to play a central role for the release of acetate. In all HDAC8-inhibitor or substrate complexes, the guanidinium group of R37 forms hydrogen bonds with the backbone carboxyl oxygen atom of two conserved glycine residues (G303 and G305) in the glycine rich loop adjacent to Y306.<sup>21–26</sup> These hydrogen bonds are formed throughout the HDAC8 simulation, however motion around the R37  $\chi_1$  angle was observed to be affected by the F152  $\chi_1$  sampling, Fig. S7.† Despite the  $\sim 10$  Å distance between F152 and R37, both residues are in proximity of the L1 loop, suggesting that the motions of F152 and R37 could be correlated with each other *via* the L1 loop.

In crystal structures of HDAC8-ligand complexes, both Y306 and F152 are present at the junction of the substrate-binding tunnel and the acetate release channel, and changes in these



two residues are important for the opening of the acetate release channel.<sup>21,28</sup> Despite this, only a weak correlation is observed between Y306  $\chi_1$  and F152  $\chi_1$  as well as between Y306 and R37, Fig. S8.† Among all the residues in the vicinity of the substrate-binding tunnel, the active site, and the product release channel,  $\chi_1$  flips are effectively only observed for the four residues described above, *i.e.*, R37, H143, F152, and Y306. For all the other residues near the active site one  $\chi_1$  conformation dominates, see Fig. S9.†

### The S39E mutation alters the sampling of both L1 and L6

The side-chain  $\chi_1$  flips of Y306, H143, and F152 described above report on the catalytic cycle of HDAC8 and the motions relate to substrate binding and product release. It is therefore of interest to investigate how these side-chain  $\chi_1$  motions are potentially altered in mutants of HDAC8 known to be substantially down-regulated. In addition to the 10  $\mu\text{s}$  simulation of wild-type HDAC8, a 10  $\mu\text{s}$  simulation was also carried out for the phosphorylation mimicking mutant S39E of HDAC8. Phosphorylation of HDAC8 at S39 as well as the S39E mutation lead to a downregulation<sup>21,53</sup> of HDAC8 enzymatic activity and since S39 is located near the L1 loop it is of interest to investigate the mechanism by which this mutation alters the dynamic sampling of the functional loops.

In the S39E-HDAC8 simulation, the dynamic sampling of Y306  $\chi_1$  and the L6 loop is inhibited and effectively only sample the *gauche*- state and the proximal conformation, respectively, Fig. 6a. Moreover, the F152  $\chi_1$  angle is highly stabilised in the *trans* state throughout the S39E simulations, which in turn leads to L1 conformations consistently distant to the active site and a very low probability of forming the K33 N<sup>z</sup>-D87 O<sup>δ</sup> salt-bridge, Fig. 6b–f. A Markov state model similar to Fig. 4c cannot be derived from the S39E-HDAC8 simulation, because the Y306(g+) state is not populated and the Y306(t) state is only present in a few frames. With regards to the kinetics observed for H143  $\chi_1$ , the H143(g-) and H143(t) states have average lifetimes of 0.023  $\mu\text{s}$  and 0.048  $\mu\text{s}$ , respectively, during the S39E-HDAC8 simulation compared to 0.020  $\mu\text{s}$  and 0.024  $\mu\text{s}$  in the simulation of wild-type HDAC8.

### The I19S CdLS mutation prevents an opening of the L6 loop

Another mutant, which has been reported in CdLS patients<sup>12</sup> and which leads to a substantial downregulation of HDAC8 is I19S. I19S-HDAC8 has an activity of <10% compared to wild-type HDAC8. The I19 residue is located at the base of the acetate release channel, Fig. 1, and about  $\sim 21$  Å from the active site. An 8  $\mu\text{s}$  MD simulation was carried out for the I19S mutant of HDAC8 (I19S-HDAC8). The first striking difference between the wild-type HDAC8 and I19S-HDAC8 simulation is that Y306, and consequently the L6 loop, are completely rigidified in the I19S-HDAC8 simulation, Fig. 7a and b. From the discussion above, this means that I19S-HDAC8 does not sample a state with an open conformation of the L6 loop, which in turn relates to product release. To rule out that the I19S-HDAC8 simulation had got trapped in a local minimum with Y306 in a *gauche*- state, the last frame of the I19S-HDAC8 simulation was taken

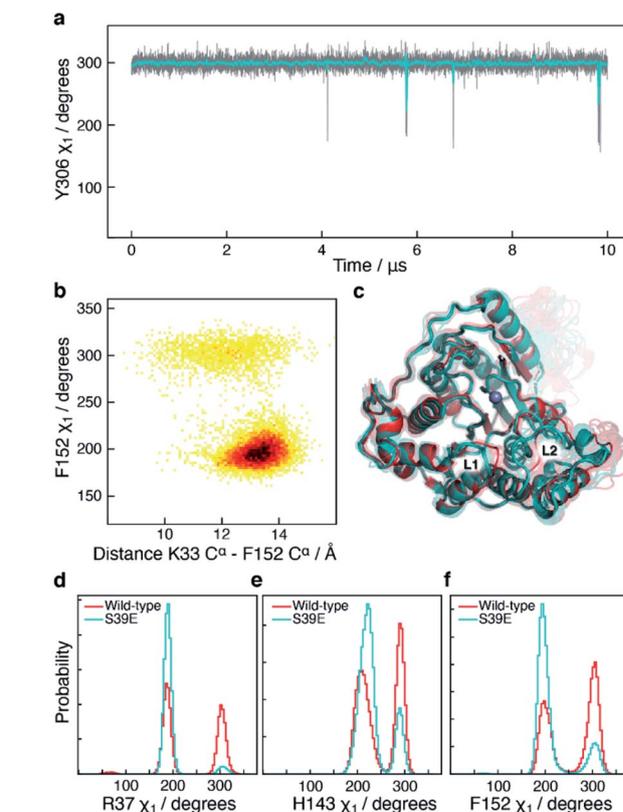


Fig. 6 Simulation of S39E-HDAC8 phosphorylation mimicking mutant. (a) The Y306  $\chi_1$  angle as a function of simulation time (grey) and sampled every nanosecond during the S39E-HDAC8 simulation. The cyan line is a moving average over a window of 20 ns. Only the Y306(g-) and the Y306(t) states are sampled, with probabilities of 99.6% and 0.4%, respectively. (b) Two-dimensional histogram depicting the correlation between the distance between K33 C $\alpha$  within the L1 loop and F152 C $\alpha$ , showing that the L1 loop is consistently distant to the substrate-binding tunnel and active site. (c) Average structure of the S39E-HDAC8 simulation in the most stable conformation, R37(t), H143(t), F152(t), Y306(g-), (cyan), along with 10 representative frames (semi-transparent cyan). Average of wild-type HDAC8 simulation in the R37(g+), H143(g+), F152(g+), Y306(g-) conformation (red), along with 10 representative frames (semi-transparent red). (d)–(f) Histograms show a substantial higher preference for R37, H143, and F152  $\chi_1$  to sample the *trans* conformation in the S39E simulation.

out and residue 19 mutated back to isoleucine, thereby recovering the wild-type sequence. A new simulation was then initiated starting from this final frame of the I19S-HDAC8 simulation, but now with residue 19 being isoleucine. After about 1  $\mu\text{s}$  of simulation time, of the back-mutated structure, Y306 starts to sample the *gauche*+ state, thus substantiating that the inhibition of the Y306 and L6 movements is caused by the I19S mutation, Fig. S10.† As for the S39E-HDAC8 simulations, a Markov state model similar to Fig. 4c cannot be derived from the I19S-HDAC8 simulation. With regards to the kinetics observed for H143  $\chi_1$ , the H143(g-) and H143(t) states have average lifetimes of 0.013  $\mu\text{s}$  and 0.009  $\mu\text{s}$ , respectively, during the I19S-HDAC8 simulation, which are significantly shorter compared to 0.020  $\mu\text{s}$  and 0.024  $\mu\text{s}$  observed in the simulation of wild-type HDAC8.



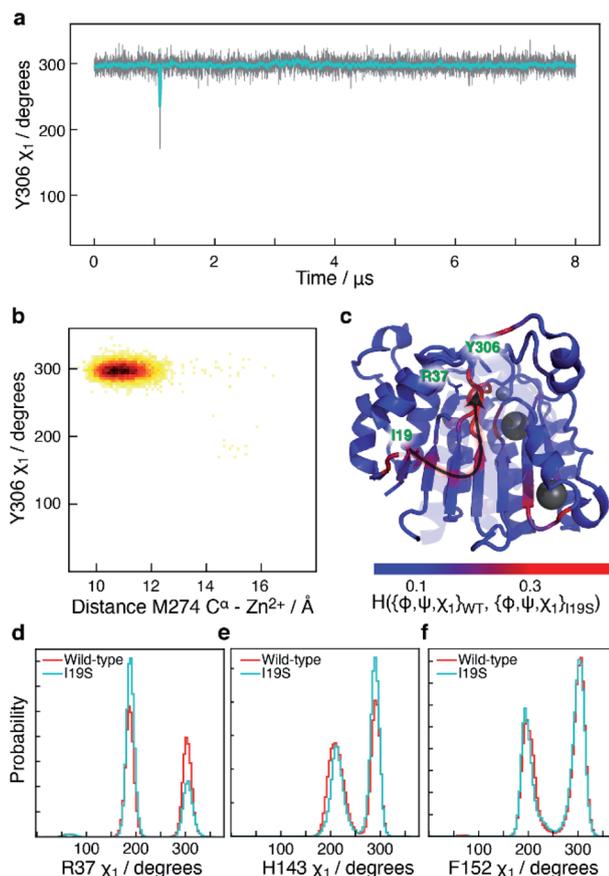


Fig. 7 Simulation of I19S-HDAC8 and comparison with wild-type. (a) The Y306  $\chi_1$  angle as a function of simulation time (grey) and sampled every nanosecond in the I19S-HDAC8 simulation. The cyan line is a moving average over a window of 20 ns. Only the Y306(g<sup>-</sup>) and the Y306(t) states are sampled, with probabilities of 99.9% and 0.1%, respectively. (b) Two-dimensional histogram showing the correlation between the distance between M274 C $\alpha$  within the L6 loop and Zn<sup>2+</sup>, and Y306  $\chi_1$ . The L6 loop only samples the conformation near the active site. (c) Residue-specific Hellinger distances between the  $\{\phi, \psi, \chi_1\}$  distribution in the I19S and wild-type simulations shown on the structure of HDAC8 (PDB: 2V5W).<sup>22</sup> A clear path between the site of mutation, I19, and Y306 is seen. (d) A slightly higher preference for a R37  $\chi_1$  *trans* conformation in the I19S simulation. (e) and (f) Very similar probability distributions in the I19S- and wild-type simulation for the H143 and F152  $\chi_1$ .

I19 and Y306 are more than 20 Å apart, both in the simulations and in all published structures of HDAC8. A central question is therefore how the I19S mutation leads to an inhibition of the  $\chi_1$  motions of Y306 as well as the motions of the L6 loop. To gain insight into the differences between the I19S and wild-type simulations, Hellinger distances were calculated between the backbone  $\{\phi, \psi\}$  and side-chain  $\chi_1$  distributions in the I19S and wild-type simulations, for each residue. A block analysis was also performed on the wild-type simulation and a Hellinger distance (I19S to wild-type) was only considered significant if it was significantly larger than what was obtained in the block analysis of the wild-type simulations (see Methods). When the significant Hellinger distances (I19S *versus* wild-type) are shown on the structure of HDAC8, Fig. 7c, a clear path is

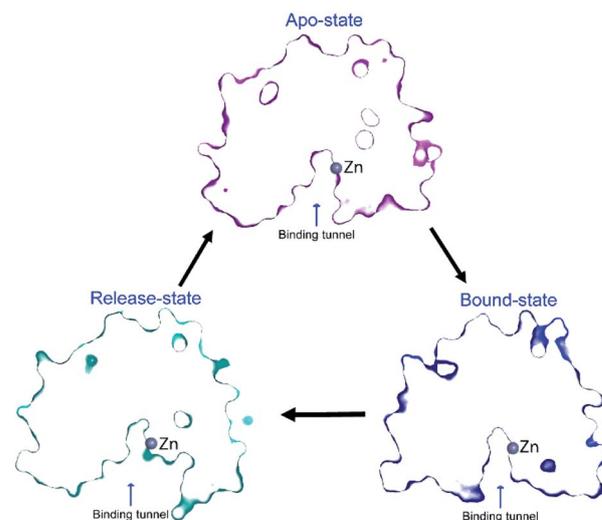


Fig. 8 Change in structure of binding tunnel of HDAC8 during catalysis. The cut-away surface views of representative structures of HDAC8 in the apo-state: {R37(g<sup>-</sup>), F152(g<sup>-</sup>), Y306(g<sup>-</sup>)} (top), release-state: {R37(t), F152(t), Y306(g<sup>+</sup>)} (bottom left), and bound-state: crystal structure 2V5W<sup>22</sup> (bottom right).

observed between residue 19 and Y306. It is important to note here that this path appears naturally and has not in any way been imposed during the calculation of the Hellinger distances. The observed path goes through the central  $\beta$ -sheet of HDAC8, including residues I135 and L299, from where the path runs *via* the glycine loop, G302–G305 to Y306. The glycine loop, G302–G305, has previously been implicated in function and mutating any of these glycine residues to alanine leads to a small change in the loop conformation in the resulting crystal structures in bound form, but causes a substantial effect on HDAC8 activity, with only about 0.3–3% of enzymatic activity retained.<sup>47</sup> Thus, not only do the Hellinger distances provide insight into how the effect of the I19S mutation is transmitted through the structure of HDAC8, but also highlight areas where mutations and regulator binding could alter HDAC8 activity due to a change in the L6 dynamics.

Arginine R37 forms interactions with the glycine loop, G302–G305. Although there is no significant change in the hydrogen-bonding between the guanidinium group of R37 and the glycine loop in the I19S-HDAC8 and wild-type HDAC8 simulations, the distribution of R37  $\chi_1$  changes slightly between the two simulations, Fig. 7d, consistent with the allosteric path going through the glycine loop. In contrast, only limited changes are observed for the distribution of the side chain  $\chi_1$  of H143 and F152 and thus of the L1 and L2 loops, Fig. 7e and f. The motion about the F152  $\chi_1$  dihedral angle is slightly faster in the I19S-simulation, where the F152(g<sup>-</sup>) has a lifetime of 1.0 ns compared to 1.4 ns in the wild-type simulation.

## Discussions

Several hypotheses have previously been suggested for the mechanism of deacetylation and downregulation of HDAC8



activity for different CdLS mutants on the basis of the structure of wild-type HDAC8 and mutants bound to various inhibitors/substrates. However, the mechanism of product release as well as the role of the dynamic loops in catalysis and how the dynamics of these are connected to key active-site residues remain elusive. Also, the effects of the various CdLS mutations and phosphorylation on the dynamics of these loops have remained largely unknown.

A substantial change in the dynamics of the L1 and L6 loops was observed upon investigation of a CdLS down-regulating mutant, I19S, as well as a phosphorylation mimicking mutation, S39E, of HDAC8. In the S39E-HDAC8 simulation all the aromatic side chains near the active site were highly stabilised in one conformation leading to a consistent open form of the L1 loop and a closed conformation of the L6 loop. Recently, Michaelis-Menten parameters,  $k_{\text{cat}}$  and  $K_{\text{M}}$ , were reported for the S39E mutant of HDAC8, where changes in both  $k_{\text{cat}}$  ( $\sim 3$  fold decrease) and  $K_{\text{M}}$  ( $\sim 2$  fold increase) were observed.<sup>53</sup> A change in  $K_{\text{M}}$  upon mutation typically reflects a change in the substrate-binding affinity, while a change in  $k_{\text{cat}}$  reflects a change in the rate of hydrolysis and/or a change in the rate of product release. These experimental data therefore suggests that mutating S39 to glutamate affects both substrate-binding as well as hydrolysis and/or product release. In the MD simulations presented above, changes in sites related to both substrate-binding (L1; F152), hydrolysis (H143), and product release (L6; Y306, M274) were observed, in line with the experimental findings.

In the simulation of I19S-HDAC8, the backbone conformation of G302–G305 changed slightly and the L6 loop as well as Y306 were trapped in a closed conformation. In available experimental enzyme kinetics data, mutation of Y306 to phenylalanine resulted in a two-fold decrease in the  $K_{\text{M}}$  and a 290-fold decrease in the catalytic constant,  $k_{\text{cat}}$ .<sup>54</sup> Similarly, upon mutation of glycine residues adjacent to Y306 to alanine, G304A and G305A, a  $\sim 300$  and a  $\sim 38$  fold decrease in  $k_{\text{cat}}$  was observed, respectively, with only a 2 fold and a 1.5 fold increase in  $K_{\text{M}}$ .<sup>47</sup> Therefore, the predominant role of Y306, whose conformational sampling is affected by the adjacent glycine residues, is in stabilising the enzymatic transition state and/or in the release of products,  $k_{\text{cat}}$ , as opposed to substrate-binding,  $K_{\text{M}}$ .<sup>54</sup> In the crystal structure of the G304A and G305A mutants, in complex with an inhibitor, only a minor change in the position of Y306 is observed relative to wild-type structures. Importantly, in these structures there are no significant change in the distance between Y306-O<sup>c</sup> and the carboxyl of the inhibitor: 2.4 Å in wild-type HDAC8 and 2.6 Å in G304A-HDAC8 and G305A-HDAC8. This strongly suggests that these mutations have limited effect on transition-state stabilisation. Hence, the glycine-to-alanine mutations likely change  $k_{\text{cat}}$  by impeding the release of product, which in turn happens *via* a change in the conformational sampling of Y306. Based on comparison of these experimental biochemical data with the results of the molecular dynamic simulations, it is reasonable to hypothesise that the downregulation caused by the I19S mutation is due to an inhibition of the L6 dynamics and a slight change in the sampling of R37, which results in the inability of I19S-HDAC8 to release the products.

During the wild-type HDAC8 simulation the Y306(g<sup>-</sup>) state, which has the substrate-binding tunnel formed, is the most stable state with an average lifetime of 1.3  $\mu\text{s}$ . Within the Y306(g<sup>-</sup>) state, R37, H143 and F152 are in rapid dynamic equilibrium between their *gauche*- and *trans* states and we therefore propose that the Y306(g<sup>-</sup>) state is a substrate-binding-competent state. In all crystal structures of HDAC8 bound to substrate or inhibitor, R37, H143, F152 and Y306 are all found in the *gauche*- state. A comparison of the HDAC8 frames with the side chain  $\chi_1$  of R37, H143, F152 and Y306 in the *gauche*- state with the crystal structures shows some differences for residues forming the active site, the substrate-binding tunnel and the release tunnel. This includes the orientation of W141 and Y306, Fig. S11.†

However, it is very likely that these side chains move slightly in the presence of substrate, to stabilise the substrate for catalysis, including forming a hydrogen-bond between Y306 and the substrate carboxyl oxygen. The Y306(g<sup>+</sup>) state can be considered as a product-release state, since the substrate-binding tunnel is widely exposed (see ESI Movie S2†). This agrees with the experimental enzyme kinetics data available for Y306F-HDAC8, where only the catalytic rate,  $k_{\text{cat}}$ , is substantially affected whereas binding of substrate is only slightly affected. A concomitant movement of R37 and F152  $\chi_1$  to *trans* in the Y306(g<sup>+</sup>) state facilitates the release of acetate. The structure of the binding tunnel in the three states: apo-state, bound-state, and release-state are shown in Fig. 8.

## Conclusions

We obtained several 10  $\mu\text{s}$ -long molecular dynamics simulations and analysed them to characterise the link between functional loops and the active site of HDAC8 as well as the effect of clinically relevant mutations on them. We observed that the conformation of aromatic side chains near the active site of HDAC8 orchestrate the sampling of the functional loops L1, L2 and L6. Specifically, side-chain movements of Y306 correlate strongly with an opening and closing of the L6 loop and side-chain  $\chi_1$  flips of F152 strongly correlate with an opening and closing of the L1 loop. The opening and closing of the functional loops relate to the catalytic cycle of HDAC8, including binding of substrate and release of products. Overall, the aromatic side-chain  $\chi_1$  movements provide the HDAC8 enzyme with a means to connect the active site with the functional loops. The presented link between functional loops and aromatic residues in the vicinity of the active site of HDAC8 provides a conceptual platform by means of which a mechanism for the regulation of HDAC8 can be derived and the mechanism of mutants from genetic disorders such as CdLS can be rationalised.

## Data availability

The trajectories associated with the research presented here are available from the corresponding author upon reasonable request.



## Author contributions

V. K. S., L. S., and D. F. H. designed the research, V. K. S. and L. S. performed molecular dynamics simulations, all authors analysed the data and molecular dynamics trajectories, and V. K. S., L. S. and D. F. H. wrote the paper. All authors discussed the results, commented on the paper, and have given approval to the final version of the manuscript. These authors contributed equally to this work: Vaibhav Kumar Shukla, Lucas Siemons.

## Conflicts of interest

The authors declare that they have no conflicts of interest with the content of this article.

## Acknowledgements

LS acknowledges the Wellcome Trust for a PhD studentship (109160/Z/15/Z). Aditi K. Borkar from the University of Nottingham as well as Edina Rosta and Martin Rosellen from University College London are acknowledged for helpful discussions and comments. The authors acknowledge the use of the UCL Myriad High Performance Computing Facility (Grace@UCL), and associated support services. This work used the ARCHER UK National Supercomputing Service (<http://www.archer.ac.uk>). This research is supported by the BBSRC (BB/R000255/1) and Leverhulme Trust (RPG-2016-268).

## References

- 1 K. Sadoul, C. Boyault, M. Pabion and S. Khochbin, *Biochimie*, 2008, **90**, 306–312.
- 2 E. Seto and M. Yoshida, *Cold Spring Harbor Perspect. Biol.*, 2014, **6**, a018713.
- 3 A. Chakrabarti, I. Oehme, O. Witt, G. Oliveira, W. Sippl, C. Romier, R. J. Pierce and M. Jung, *Trends Pharmacol. Sci.*, 2015, **36**, 481–492.
- 4 M. Haberland, R. L. Montgomery and E. N. Olson, *Nat. Rev. Genet.*, 2009, **10**, 32–42.
- 5 S. Minucci and P. G. Pelicci, *Nat. Rev. Cancer*, 2006, **6**, 38–51.
- 6 D.-M. Chuang, Y. Leng, Z. Marinova, H.-J. Kim and C.-T. Chiu, *Trends Neurosci.*, 2009, **32**, 591–601.
- 7 M. A. Deardorff, M. Bando, R. Nakato, E. Watrin, T. Itoh, M. Minamino, K. Saitoh, M. Komata, Y. Katou, D. Clark, K. E. Cole, E. De Baere, C. Decroos, N. Di Donato, S. Ernst, L. J. Francey, Y. Gyftodimou, K. Hirashima, M. Hullings, Y. Ishikawa, C. Jaulin, M. Kaur, T. Kiyono, P. M. Lombardi, L. Magnaghi-Jaulin, G. R. Mortier, N. Nozaki, M. B. Petersen, H. Seimiya, V. M. Siu, Y. Suzuki, K. Takagaki, J. J. Wilde, P. J. Willems, C. Prigent, G. Gillessen-Kaesbach, D. W. Christianson, F. J. Kaiser, L. G. Jackson, T. Hirota, I. D. Krantz and K. Shirahige, *Nature*, 2012, **489**, 313–317.
- 8 X. Gao, Z. Huang, Y. Fan, Y. Sun, H. Liu, L. Wang, X.-F. Gu and Y. Yu, *Cell. Physiol. Biochem.*, 2018, **47**, 2388–2395.
- 9 A. A. Lane and B. A. Chabner, *J. Clin. Oncol.*, 2009, **27**, 5459–5468.
- 10 M. Dokmanovic, C. Clarke and P. A. Marks, *Mol. Cancer Res.*, 2007, **5**, 981–989.
- 11 T. Eckschlager, J. Plch, M. Stiborova and J. Hrabeta, *Int. J. Mol. Sci.*, 2017, **18**, 1414.
- 12 F. J. Kaiser, M. Ansari, D. Braunholz, M. Concepción Gil-Rodríguez, C. Decroos, J. J. Wilde, C. T. Fincher, M. Kaur, M. Bando, D. J. Amor, P. S. Atwal, M. Bahlo, C. M. Bowman, J. J. Bradley, H. G. Brunner, D. Clark, M. Del Campo, N. Di Donato, P. Diakumis, H. Dubbs, D. A. Dymont, J. Eckhold, S. Ernst, J. C. Ferreira, L. J. Francey, U. Gehlken, E. Guillén-Navarro, Y. Gyftodimou, B. D. Hall, R. Hennekam, L. Hudgins, M. Hullings, J. M. Hunter, H. Yntema, A. M. Innes, A. D. Kline, Z. Krumina, H. Lee, K. Leppig, S. A. Lynch, M. B. Mallozzi, L. Mannini, S. Mckee, S. G. Mehta, I. Micule, S. Mohammed, E. Moran, G. R. Mortier, J.-A. S. Moser, S. E. Noon, N. Nozaki, L. Nunes, J. G. Pappas, L. S. Penney, A. Pérez-Aytés, M. B. Petersen, B. Puisac, N. Revencu, E. Roeder, S. Saitta, A. E. Scheuerle, K. L. Schindeler, V. M. Siu, Z. Stark, S. P. Strom, H. Thiese, I. Vater, P. Willems, K. Williamson, L. C. Wilson, H. Hakonarson, F. Quintero-Rivera, J. Wierzba, A. Musio, G. Gillessen-Kaesbach, F. J. Ramos, L. G. Jackson, K. Shirahige, J. Pié, D. W. Christianson, I. D. Krantz, D. R. Fitzpatrick and M. A. Deardorff, *Hum. Mol. Genet.*, 2014, **23**, 2888–2900.
- 13 C. Decroos, C. M. Bowman, J.-A. S. Moser, K. E. Christianson, M. A. Deardorff and D. W. Christianson, *ACS Chem. Biol.*, 2014, **9**, 2157–2164.
- 14 C. Decroos, N. H. Christianson, L. E. Gullett, C. M. Bowman, K. E. Christianson, M. A. Deardorff and D. W. Christianson, *Biochemistry*, 2015, **54**, 6501–6513.
- 15 L. Feng, D. Zhou, Z. Zhang, Y. Liu and Y. Yang, *J. Hum. Genet.*, 2014, **59**, 536–539.
- 16 T. Saikusa, M. Hara, K. Iwama, K. Yuge, C. Ohba, J. Okada, T. Hisano, Y. Yamashita, N. Okamoto, H. Saitsu, N. Matsumoto and T. Matsuishi, *Brain Dev.*, 2018, **40**, 406–409.
- 17 N. D. Werbeck, V. K. Shukla, M. B. A. Kunze, H. Yalinca, R. B. Pritchard, L. Siemons, S. Mondal, S. O. R. Greenwood, J. Kirkpatrick, C. M. Marson and D. F. Hansen, *Nat. Commun.*, 2020, **11**, 3841.
- 18 P. J. Watson, L. Fairall, G. M. Santos and J. W. R. Schwabe, *Nature*, 2012, **481**, 335–340.
- 19 C. J. Millard, P. J. Watson, I. Celardo, Y. Gordiyenko, S. M. Cowley, C. V. Robinson, L. Fairall and J. W. R. Schwabe, *Mol. Cell*, 2013, **51**, 57–67.
- 20 J. D. Osko, N. J. Porter, C. Decroos, M. S. Lee, P. R. Watson, S. E. Raible, I. D. Krantz, M. A. Deardorff and D. W. Christianson, *J. Struct. Biol.*, 2021, **213**, 107681.
- 21 J. R. Somoza, R. J. Skene, B. A. Katz, C. Mol, J. D. Ho, A. J. Jennings, C. Luong, A. Arvai, J. J. Buggy, E. Chi, J. Tang, B.-C. Sang, E. Verner, R. Wynands, E. M. Leahy, D. R. Dougan, G. Snell, M. Navre, M. W. Knuth,



- R. V. Swanson, D. E. McRee and L. W. Tari, *Structure*, 2004, **12**, 1325–1334.
- 22 A. Vannini, C. Volpari, P. Gallinari, P. Jones, M. Mattu, A. Carfí, R. De Francesco, C. Steinkühler and S. Di Marco, *EMBO Rep.*, 2007, **8**, 879–884.
- 23 M. A. Holbert and R. Marmorstein, *Curr. Opin. Struct. Biol.*, 2005, **15**, 673–680.
- 24 A. Vannini, C. Volpari, G. Filocamo, E. C. Casavola, M. Brunetti, D. Renzoni, P. Chakravarty, C. Paolini, R. De Francesco, P. Gallinari, C. Steinkühler and S. Di Marco, *Proc. Natl. Acad. Sci. U. S. A.*, 2004, **101**, 15064–15069.
- 25 K. E. Cole, D. P. Dowling, M. A. Boone, A. J. Phillips and D. W. Christianson, *J. Am. Chem. Soc.*, 2011, **133**, 12474–12477.
- 26 A. A. Tabackman, R. Frankson, E. S. Marsan, K. Perry and K. E. Cole, *J. Struct. Biol.*, 2016, **195**, 373–378.
- 27 M. A. Deardorff, N. J. Porter and D. W. Christianson, *Protein Sci.*, 2016, **25**, 1965–1976.
- 28 L. Whitehead, M. R. Dobler, B. Radetich, Y. Zhu, P. W. Atadja, T. Claiborne, J. E. Grob, A. McRiner, M. R. Pancost, A. Patnaik, W. Shao, M. Shultz, R. Tichkule, R. A. Tommasi, B. Vash, P. Wang and T. Stams, *Bioorg. Med. Chem.*, 2011, **19**, 4626–4634.
- 29 S. L. Gantt, C. G. Joseph and C. A. Fierke, *J. Biol. Chem.*, 2010, **285**, 6036–6043.
- 30 R. Wu, S. Wang, N. Zhou, Z. Cao and Y. Zhang, *J. Am. Chem. Soc.*, 2010, **132**, 9471–9479.
- 31 N. D. Werbeck, J. Kirkpatrick, J. Reinstein and D. F. Hansen, *ChemBioChem*, 2014, **15**, 543–548.
- 32 K. Chen, X. Zhang, Y.-D. Wu and O. Wiest, *J. Am. Chem. Soc.*, 2014, **136**, 11636–11643.
- 33 B. Webb and A. Sali, *Curr. Protoc. Bioinf.*, 2016, **54**, 5.6.1–5.6.37.
- 34 V. B. Chen, W. B. Arendall, J. J. Headd, D. A. Keedy, R. M. Immormino, G. J. Kapral, L. W. Murray, J. S. Richardson and D. C. Richardson, *Acta Crystallogr., Sect. D: Biol. Crystallogr.*, 2010, **66**, 12–21.
- 35 W. L. Jorgensen, J. Chandrasekhar, J. D. Madura, R. W. Impey and M. L. Klein, *J. Chem. Phys.*, 1983, **79**, 926–935.
- 36 K. Lindorff-Larsen, S. Piana, K. Palmo, P. Maragakis, J. L. Klepeis, R. O. Dror and D. E. Shaw, *Proteins: Struct., Funct., Bioinf.*, 2010, **78**, 1950–1958.
- 37 H. J. C. Berendsen, J. P. M. Postma, W. F. van Gunsteren, A. DiNola and J. R. Haak, *J. Chem. Phys.*, 1984, **81**, 3684–3690.
- 38 M. Parrinello and A. Rahman, *J. Appl. Phys.*, 1981, **52**, 7182–7190.
- 39 S. Nosé, *J. Chem. Phys.*, 1984, **81**, 511–519.
- 40 W. G. Hoover, *Phys. Rev. A: At., Mol., Opt. Phys.*, 1985, **31**, 1695–1697.
- 41 G. Bussi, D. Donadio and M. Parrinello, *J. Chem. Phys.*, 2007, **126**, 014101.
- 42 B. Hess, H. Bekker, H. J. C. Berendsen and J. G. E. M. Fraaije, *J. Comput. Chem.*, 1997, **18**, 1463–1472.
- 43 U. Essmann, L. Perera, M. L. Berkowitz, T. Darden, H. Lee and L. G. Pedersen, *J. Chem. Phys.*, 1995, **103**, 8577–8593.
- 44 M. J. Abraham, T. Murtola, R. Schulz, S. Páll, J. C. Smith, B. Hess and E. Lindahl, *SoftwareX*, 2015, **1–2**, 19–25.
- 45 G. R. Bowman, V. S. Pande and F. Noé, *Adv. Exp. Med. Biol.*, 2014, **797**, 7–22.
- 46 N. Michaud-Agrawal, E. J. Denning, T. B. Woolf and O. Beckstein, *J. Comput. Chem.*, 2011, **32**(10), 2319–2327.
- 47 N. J. Porter, N. H. Christianson, C. Decroos and D. W. Christianson, *Biochemistry*, 2016, **55**, 6718–6729.
- 48 P. M. Lombardi, H. D. Angell, D. A. Whittington, E. F. Flynn, K. R. Rajashankar and D. W. Christianson, *Biochemistry*, 2011, **50**, 1808–1817.
- 49 M. J. Bottomley, P. Lo Surdo, P. Di Giovine, A. Cirillo, R. Scarpelli, F. Ferrigno, P. Jones, P. Neddermann, R. De Francesco, C. Steinkühler, P. Gallinari and A. Carfí, *J. Biol. Chem.*, 2008, **283**, 26694–26704.
- 50 M. B. A. Kunze, D. W. Wright, N. D. Werbeck, J. Kirkpatrick, P. V. Coveney and D. F. Hansen, *J. Am. Chem. Soc.*, 2013, **135**, 17862–17868.
- 51 D. P. Dowling, S. L. Gantt, S. G. Gattis, C. A. Fierke and D. W. Christianson, *Biochemistry*, 2008, **47**, 13554–13563.
- 52 M. Marek, S. Kannan, A.-T. Hauser, M. Moraes Mourão, S. Caby, V. Cura, D. A. Stolfa, K. Schmidt-kunz, J. Lancelot, L. Andrade, J.-P. Renaud, G. Oliveira, W. Sippl, M. Jung, J. Cavarelli, R. J. Pierce and C. Romier, *PLoS Pathog.*, 2013, **9**, e1003645.
- 53 K. R. Welker Leng, C. A. Castañeda, C. Decroos, B. Islam, S. M. Haider, D. W. Christianson and C. A. Fierke, *Biochemistry*, 2019, **58**, 4480–4493.
- 54 S. M. L. Gantt, C. Decroos, M. S. Lee, L. E. Gullett, C. M. Bowman, D. W. Christianson and C. A. Fierke, *Biochemistry*, 2016, **55**, 820–832.

