Meaningful machine learning models and machine-learned pharmacophores from fragment screening campaigns

Carl Poelking,1,2, ∗ Gianni Chessari,2 Christopher W. Murray,2 Richard J. Hall,2 Lucy Colwell,1 and Marcel L. Verdonk2

1Department of Chemistry, University of Cambridge, UK
2Astex Pharmaceuticals, Cambridge, UK

Machine learning (ML) is widely used in drug discovery to train models that predict protein-ligand binding. These models are of great value to medicinal chemists, in particular if they provide case-specific insight into the physical interactions that drive the binding process. In this study we derive ML models from over 50 fragment-screening campaigns to introduce two important elements that we believe are absent in most – if not all – ML studies of this type reported to date: First, alongside the observed hits we use to train our models, we incorporate true misses and show that these experimentally validated negative data are of significant importance to the quality of the derived models. Second, we provide a physically interpretable and verifiable representation of what the ML model considers important for successful binding. This representation is derived from a straightforward attribution procedure that explains the prediction in terms of the (inter-)action of chemical environments. Critically, we validate the attribution outcome on a large scale against prior annotations made independently by expert molecular modellers. We find good agreement between the key molecular substructures proposed by the ML model and those assigned manually, even when the model’s performance in discriminating hits from misses is far from perfect. By projecting the attribution onto predefined interaction prototypes (pharmacophores), we show that ML allows us to formulate simple rules for what drives fragment binding against a target automatically from screening data.

Fragment-based drug discovery (FBDD) is now a well-established method for obtaining weak, low-molecular weight hits that provide efficient starting points for hit-to-lead optimisation against a protein target of interest [1, 2]. The output from fragment screening campaigns provides a rich source of data for understanding the types of interactions that drive protein-ligand binding for a given protein. Firstly, because of their small size and low complexity, fragment screens tend to provide significantly higher hit rates than screening campaigns involving larger molecules [3]. Additionally, particularly for well-designed fragment libraries, differences between hits (actives) and misses (inactives) in terms of functional groups or physicochemical profile should be easier to identify than for libraries containing larger, more decorated compounds. However, manually analysing and interpreting the output from a fragment screening campaign of 1000 fragments can be challenging. The current study shows how the careful use of machine-learning (ML) methods can automate this process.

ML approaches have routinely and successfully assisted the drug discovery workflow for many decades now [4]. Over recent years, the rapid evolution and adoption of ML has contributed a large array of novel techniques and ideas [5, 6]: These include new strategies for modelling of chemically sparse data [7], deep representations of molecular structures [8], neural-network based docking and scoring functions [9] coupled with data augmentation [10], data-driven predictions of force-field parameters [11], generative modelling [12], imputation of assay data [13], etc. Some of these strategies are, however, not without problems [14, 15]. Indeed, there are perhaps three reasons why the adoption of these exciting technologies may lag behind the astonishing pace at which new techniques are being developed: First, a chronic lack of reliable training data. Second, a chronic lack of reliable models. Third, a chronic lack of trust in the reliability of the (reliable) models. This lack of trust is partly blamed on the black-box character of any sufficiently sophisticated machine-learning framework – which, as a cliché, is partly true, and partly outdated: Tremendous effort has been made over recent years to improve the interpretability of machine-learned models through improved attribution techniques and visualization [16].

In a drug discovery context, attribution involves mapping the predicted output (such as a binding affinity) back onto features of the input (the molecular structure). The complexity of the attribution procedure grows with the complexity of the underlying machine-learning model. Still, even for simple molecular-fingerprint-based methods, deciding on the quality of this attribution can be a contentious issue, as disagreement between models is commonplace, and the “true” answer to the attribution problem is typically unknown [15, 17]. In particular, Sheridan observed that standard ML architectures (in his case, different combinations of fingerprints and classifiers) routinely disagree in their attributions, despite agreeing with respect to the predicted activities [17]. Using synthetic data, McCloskey et al. furthermore noted that ML classifiers may exist in a state of deceptive bliss, where the predictions as measured by standard metrics appear close to perfect, but the inferred binding logic has severe deficiencies [15]. Finally, Sundar et al. proposed decoys as a means to avoid attribution false negatives, and ascribed attribution false positives to background correlations caused by finite sampling. Still, in the development and validation of new attribution procedures, these studies share a common hurdle – namely, the lack of a ground truth derived from realistic experimental rather than synthetic data. Being able to assess and compare models based on this more fine-grained ground truth rather than on binary binding labels alone would be of great benefit by enforcing chemical realism and thus reducing false positives in virtual

∗Electronic address: carl.poelking@astx.com
screening campaigns [18].

For FBDD in particular, interpretation through attribution is desirable because it enables medicinal chemists to make sense of the output of a fragment-screening campaign; pointing, for example, to a specific set of interactions that need to be preserved for binding to occur. In this paper we show that identifying the "hot" regions within fragment hits can be achieved remarkably well with ML methods if sensible attribution strategies are paired with carefully balanced datasets incorporating "true" hits and "true" misses. We use data from over 50 fragment screening campaigns, both from internal Astex projects as well as from fragment screens performed at the Diamond Light Source to train ML models designed to distinguish hits from misses. For the Astex internal datasets the misses are experimentally validated inactives, which as we illustrate enhances the ML method's ability to provide accurate interpretable models. The attribution procedure becomes robust and virtually unambiguous by resorting to an atomic-environment-based description of the molecular structure, kernel learning and simple filtering. We thus identify which regions in the fragments drive binding against each target. Crucially, we verify that the attribution of the ML binding models is not only consistent internally, but also with expert judgement. The validation is thus approached from two angles using, first, a metric for the spatial co-localization of attribution weights in the context of the binding site; second, a large number of manual annotations that preceded the modelling. The analysis shows that the ML models are able to successfully identify the "hot" regions within fragments across a wide range of targets.

Results & Discussion

Datasets

The fragment binding data we use here stems from two sources: The Astex dataset, encompassing 52 different protein targets from Astex projects; and the XChem dataset, consisting of 26 targets [19, 20]. Both datasets are derived from the
Figure 2: AUC statistics aggregating the results from the Astex and XChem data for the (a) site-specific models, (b) kinase model, (c) global model. The red bars indicate the standard error of the AUC for each test case. The blue bars in (b) and (c) correspond to AUCs measured for the input labels $y_{hit}^A$ rather than predicted scores $z_{hit}^A$ (see main text for discussion). Lighter shading indicates results obtained for the XChem set. The pocket index enumerates the different pockets across all targets for which independent models were trained, sorted here by the test prediction AUC. For the kinase model, only the pockets corresponding to the active site are considered. (d) Autocorrelation metrics $\bar{\phi}_\pm$ capturing non-random spatial correlations of the attribution weights across the superimposed fragment hit densities. Red, grey and blue bars correspond to: $\bar{\phi}_+$ in decreasing order, $\bar{\phi}_-$ in increasing order, and the null metric $\bar{\phi}_{0+}$, respectively. (e) Autocorrelation metrics for models derived from validated hits augmented with synthetic decoys (saturated colours) compared to models derived from validated hits and misses (pale curves, as reproduced from panel (d) for systems where validated misses were available). Note the grey bars in panels (d) and (e): These indicate the AUCs of the pockets when sorted in accordance with the co-localization metric of the positive (red) branch, and illustrating the vanishing correlation between the prediction and attribution performance.

direct use of X-ray crystallography in a screening mode where a library of diverse fragments were either soaked into protein crystals or co-crystallised with the protein. Within the Astex data, a fragment is designated a “validated hit” only if clear and unambiguous electron density could be assigned to it. For the analysis we define as the Astex fragment library the union of all the fragments that at some stage formed part of the (evolving) Astex X-ray fragment library, whereas the XChem fragment library was approximated via the union of all reported fragment hits across the different screening campaigns (see Methods section). After postprocessing we obtain 1660 validated hits (26900 misses) for the Astex, and 690 validated hits (18200 misses) for the XChem dataset. Next to site-specific models that consider individual binding pockets or allosteric sites on a protein target, we also construct global hit-rate and kinase hit-rate models that are trained on data aggregated across a larger set of proteins (see Methods section for details).

Fragments are typically selected to be small yet diverse, contain functional groups representative of drug-like compounds, have high aqueous solubility and desirable physical properties such as low lipophilicity. To build intuition for the library composition and the relative frequency of different chemical environments, we include in Fig. 1 visualizations of the physicochemical and structural space covered by the fragment libraries. The distributions of heavy-atom counts, rotatable bonds and ClogP (calculated logP, a measure of lipophilicity) in Fig. 1a-c indicate that the XChem fragment library involves somewhat larger and more complex molecules than the Astex fragment library (red histograms). The distribution of pairwise similarities $k_{ab} = \langle x_a, x_b \rangle$ in Fig. 1d ($x_a, x_b$ are descriptor vectors of atomic environments $a, b$, see Methods) highlights the large diversity of both and, in particular, the Astex set on the sub-fragment level: Fig. 1e elaborates on this further using a low-dimensional projection of the atomic environments that are part of the Astex fragment library, obtained by approximately reproducing the distance $d_{ab}$ in the 2D plane using a harmonic-network optimization. Each
point of the projection represents an atom-centred chemical environment and is coloured according to the atomic element of that atom, with the lines in this plot representing covalent bonds between atoms. Some major clusters can be assigned to: aliphatic and aromatic carbon environments (centre left and centre right), carbon in aromatic heterocycles (centre), amines, amidines and conjugated nitrogen (bottom left), and acids (top left). We will return to these maps to elucidate the activity of chemical environments on a coarse-grained level.

### Fragment binding models

We briefly summarize the key outputs of our ML approach (see the Methods section for details). Once trained on fragment-binding data of validated hits and misses, our ML model equips us with: a local environment-based and global molecular similarity measure (kernel); an overall binding score \( Z_A = \sum_{a \in A} z_a \) for a molecule \( A \) with atomic-environment-based attributions \( z_a \); and a predicted overall binding label \( y_A = \text{sign}(Z_A) \) (to be compared against the true experimental label \( y_A^{\text{true}} \)).

One of the aims of this paper is to rigorously assess the physical realism of the atom-centred attribution weights \( z_a \). First, however, we verify that the individual binding models are able to adequately classify test compounds into hits and misses. We consider three different model types: Site models, which are derived for a particular pocket on a particular target; kinase (class) models, trained by aggregating the sets of hits and misses across the active sites of multiple kinase targets; and global models, trained by aggregating hits and misses across multiple targets and pockets, whatever their type or identity.

The classification results from the cross-validation experiments are summarized in Fig. 2. The testing protocol for the site-specific models is based on random cross-validation (CV) with a training fraction of 0.7: This validation mode can be problematic if the set of hits includes many close analogues of the same warhead, hence exposing the models to structural bias. Fragment libraries, however, are designed around low-mW compounds sampled from a diverse, unbiased set. Even simple CV thus gives an adequate estimate of a model’s predictive power. We use the area under the receiver operating characteristic (AUC) as performance metric. The results for the individual pockets from the Astex set (see Fig. 2a, dark colours) range between AUCs of 0.92 down to 0.55. We note that some of the models trained on the XChem data (light colours) achieve higher AUCs, due to the less challenging composition of those datasets (which can include analogues) and lack of validated misses. The average AUC of around 0.75 is expectedly low given the challenging composition (and, interestingly, increases considerably as true negatives are replaced by decoys, see below).

For the kinase class model (Fig. 2b), we adapt the CV procedure to mimic more closely the situation of a prospective screen against a new kinase target: We derive the classification of fragments into low- vs high-hit-rate compounds from all but one of the kinase datasets and then train a model on these data. Finally we use the hit-rate labels \( \{y_A^h\} \) (separating high \( = +1 \) from low \( = -1 \) hit rates, see Methods) and hit-rate scores \( \{Z_A^h\} \) predicted for the kinase left out in the training as a proxy for the binding labels \( \{y_A\} \) and scores \( \{Z_A\} \), respectively. The resulting AUCs, with each kinase target left out in turn, are shown in Fig. 2b (red bars): They are surprisingly high, with all but two systems achieving a value larger than 0.75. Intriguingly, the AUCs obtained by ranking the compounds by the predicted scores \( \{Z_A^h\} \) are better than if we rank them by the true hit-rate labels \( \{y_A^h\} \) on which the models were trained (blue bars). This is a first indicator that the machine-learning model is able to identify relevant sub-patterns within the molecular structures to a degree that goes beyond global molecular recognition or scaffold recall [14]. Similar conclusions hold for the global model (Fig. 2c), which is evaluated analogously.

Returning to the site-specific models, we note that we also trained baseline models derived from topological molecular fingerprints used in either a kernel setting or logistic regression (see Methods for details). Whereas the topological kernel has a slight edge over the logistic regression, the measured AUCs are nevertheless on average by 0.05 (or approximately 8%) lower than those achieved by the convolutional best-match \( G_m \cdot Y_{\text{tm}} \) kernel described in the Methods section. The latter is therefore attractive both regarding the prediction performance as well as the uniqueness and ease of the attribution procedure. This procedure is somewhat more involved and requires, e.g., iterative dropout applied to molecular substructures and/or input features when global topological fingerprints are paired with nonlinear classifiers such as neural networks or logistic layers [17, 21].

### Attribution validation

The aim of this section is to achieve a statistical validation of the attribution procedure that goes beyond an anecdotal case-by-case confirmation of individual attribution outcomes.

By projecting the whitened (i.e., centered and scaled) atomic attribution weights \( \tilde{z}_a \) (see Methods, Eq. 7) onto the chemical maps in Fig. 1, we gain a first coarse-grained understanding of which regions of chemical space a model looks at most in arriving at its predictions. For the global and kinase models (Fig. 1f and g) a fairly clear picture results. Globally (and expectedly) we observe a strong lipophilic hotspot (in black) corresponding to aromatic carbon environments (centre right) and sulphurs (top right). Aliphatic carbons on the other hand remain almost surprisingly quiet: Notice the absence of the attribution signal on the left-hand side of the carbon domain, which also reflects the higher hit-rate achieved by flat structures. For kinases, nitrogen groups (in particular, aromatic nitrogen, center bottom) dominate. Finally, for the site-specific model (Fig. 1h, corresponding to the active site of the serine protease uPA in this case), the activity cannot be reduced to a single chemical cluster. The attribution, however, still points to a simple underlying pharmacophore, as exemplified by the highlighted amino-isoquinoline.

On a case-by-case basis such a more detailed picture
Figure 3: Comparison between manual and machine-learned annotations of superimposed fragment hits for eight systems (a-h). Each triptych consists of: left, the superimposed ligand density; centre, the manually assigned pharmacophore weights mapped onto this density; right, the attribution field derived from the filtered (ranked) weights. The superposition was obtained by aligning the binding sites of the X-ray structures of the protein-ligand complexes. All densities are visualised on the plane of best fit corresponding to the first two principal components of the nuclear coordinates.

is arrived at by projecting the attribution weights $z_a$ onto individual molecular structures. Prior to this projection, an additional filtering step maps $z_a$ onto an attribution signal $F(z_a) \in [0, 1]$, with larger values of $F(z)$ indicating higher confidence that the assigned weight is statistically significant. In Fig. 1h, we thus see for a particular fragment hit against uPA how the model highlights a feature on the 2-aminopyridine substructure, which, indeed, forms a salt bridge with the protein. For the kinase model (panel g), the pyrazole of a phenylpyrazole derivative is highlighted, again perfectly in line with chemical intuition. For the global model (panel f) we consider the same isoquinoline derivative as tested against uPA: this time, however, the lipophilic region complementary to the salt-bridge-forming amidine lights up, reproducing the fact that in general lipophiles tend to achieve higher hit rates.

Co-localization metric. For a first more comprehensive validation of our attribution procedure, we use the relative positioning of the ligands within their binding site as derived from spatially aligned X-ray structures: The underlying assumption is that hotspots on the ligands correspond to hotspots on the protein. Environments with large attribution weights, when mapped onto the superimposed X-ray binding modes of the fragments, should thus cluster (i.e., “co-localize”) in space. This information – which is not made available to the ML models during training – leads to a superposition density onto which we can subsequently project the attribution weights. The validation step then consists of evaluating a spatial autocorrelation statistic $|\bar{\varphi}_z| \leq 1$ of the attribution weights relative to their local attribution “field” (see Methods for the definition of $\bar{\varphi}_z$). An autocorrelation with a magnitude larger than a random baseline indicates that the attribution procedure was
successful in the sense that chemical environments deemed significant in the binding tend to localize in the same region of the binding site.

Fig. 2d summarizes the co-localization metrics $\phi_{\pm}$ calculated for a range of binding sites. To distinguish between the prediction of “hot” and “cold” spots, $\phi_+$, indicated by the red bars, measures the co-localization of chemical environments with positive (higher than average) weights; $\phi_-$, indicated by the grey bars, measures co-localization of environments with negative (lower than average) weights. The blue bars indicate the null background $\phi_0$ derived from scrambled (randomly permuted) weights. The comparison with this null background highlights that the attribution weights exhibit significant non-random spatial clustering. The autocorrelation of the weights is more pronounced for the positive branch $\phi_+$ than the negative branch $\phi_-$, indicating that the hotspots do indeed cluster significantly better in space than the predicted cold spots. This is not surprising given that the autocorrelations are computed only across fragment hits, which as such exclude clashes that could result in large negative attribution weights.

To understand better the role played by the inactive data (i.e., validated misses) in informing the attribution model, we repeat this analysis for synthetic datasets that combine validated hits from the Astex sets (for which we have true validated misses) with unvalidated decoys sampled from the XChem fragment library. As shown in Fig. 2e, this replacement of misses with decoys results in a noticeable decay of the autocorrelation signal, whereas the prediction AUCs increase drastically to almost unity (see the pale grey bars at the top of this plot compared to those in panel d): This shows that the models get overconfident in forming their predictions, while the attribution becomes less specific and smeared out – a finding that agrees well with previous studies that observed that large AUCs are not necessarily a good indicator of chemical soundness [14, 15]. Even for the original datasets that incorporate both validated hits and misses, the correlation between the autocorrelation metric and AUCs is marginal at best (see the red bars and pale grey bars at the top of Fig. 2d, respectively). This disconnect between classification and attribution performance reminds us that we are still in a sampling regime where the inferred fragment logic does not have to be complete to arrive at (coincidentally) correct predictions. True negative data then acts as a natural filter that forces the models to assign atomically “sparse” and meaningful attributions.

**Expert annotations.** The co-localization analysis indicates that the attribution procedure is internally consistent, with hotspots inferred by the machine-learning protocol clustering together within the binding site. We now complement this validation route with a yet more stringent test, comparing the derived hotspots with those annotated manually by human experts. These annotations, performed prior to any machine learning, involved assigning a minimal pharmacophore to each fragment hit by visually inspecting the fragment-protein complex. The minimal pharmacophore of a fragment hit reflects what interaction motif drives the binding of the fragment to the protein and which atoms in the fragment are involved in these interactions. Examples of minimal pharmacophores include a donor-acceptor pattern in the fragment forming a pair of hydrogen bonds with the hinge region of a kinase target, or an amidine group forming an ion pair with the aspartic acid in the S1 pocket of a serine protease. The definitions of these minimal pharmacophores are stored as SMARTS patterns that are mapped onto the chemical structures. We denote this mapping as a spin vector for each molecule with components $m_a = \pm 1$ that indicate whether an atom $a \in A$ forms part of the pharmacophore $(m_a = +1)$ or not $(m_a = -1)$.

To assess the degree to which the annotations $\{m_a\}$ agree with our attribution weights, we map both onto the superimposed point densities of the fragments and project the resulting fields onto the plane of best fit of the ligand density. Fig. 3 shows a visual comparison of heat map representations of these fields for a subset of eight binding sites, juxtaposed with the superimposed densities. For a quantitative comparison, we evaluate the Pearson correlation $\rho$ between the attribution and annotation fields. This correlation is surprisingly large, ranging between $\rho = 0.58$ to $\rho = 0.95$, and, again, only weakly correlated with the classification metrics achieved by the machine learning: E.g., the site model with the largest observed $\rho = 0.95$ has an AUC of 0.77, whereas the uPA model achieved an AUC of 0.89, but a field correlation $\rho$ of only 0.74. This can be rationalized in that simple, but potentially non-specific pharmacophores (such as an acid group) tend to lead to good agreement between the manual and automatic annotation, but poor classification. More precisely, due to the low specificity of the pharmacophore, but also the difficulty of predicting steric clashes (among others), the model has trouble resolving why some fragments featuring the same pharmacophore were reported as a validated miss. The human experts on the other hand may not face the same conundrum, as their annotations are not always (and not easily) checked manually for consistency with the set of misses. This in turn offers another perspective how the attribution output can be combined with manual annotations to check for weakly specific or incomplete pharmacophore assignments. Nevertheless, already in the present form, both the quantitative and visual comparison indicate widespread agreement between the machine-learned and manual attributions both for strongly localized, simple pharmacophores (such as the acid group in Fig. 3g) as well as more diffuse pharmacophores (such as those of Fig. 3f).

A potentially problematic issue concerns the degree to which the machine-learned attribution localizes the inferred binding pattern within a given fragment hit. This issue is partly connected to how we originally defined the chemical environments: They are atom-centred, and have a size that is determined by a radial cutoff of, here, 5.5 Å. This implies that, given a large attribution weight, in principle any structural feature contained within the cutoff sphere of an atom could be deemed significant. Fortunately, in practice the attribution tends to localize important groups with a higher resolution than this cutoff radius. We attribute this to the concerted action of the smooth atomic kernel, the competitive matching procedure performed on the molecular level, and post-processing of the weights. Amplified further by effective decoys (ideally, validated misses), the localization can be remarkably strong – see,
PredictionAnnotation
a b c
fs fs fs

figure T: projections of machine-learned and manually assigned pharmacophore amplitudes onto two-dimensional pharmacophore maps derived from co-occurrence statistics. The projections are obtained from the site-specific models for (a) the ATP site of the kinase CDK2 (b) the protein-protein interaction target KEAP1, (c) the trypsin-like serine protease uPA. The colour maps for the top row indicate the locally averaged pharmacophore amplitude $\hat{A}$ (Eq. 11). For the bottom row, the colour reflects the relative frequency of a pharmacophore across the set of annotated fragment hits.

for example, the large weight attributed to the covalently bound acetyl group in the case of the SARS-CoV-2 main protease (see Fig. S2 of the SI appendix) [22].

Pharmacophore projections

The analysis of the machine-learned attribution fields illustrates that the attribution procedure often works remarkably well. We now address, as a perspective, whether we can derive explicit (human-readable) pharmacophore trends that go beyond case-by-case visualizations such as those shown in Fig. 1f-h. The idea is to map the attribution weights onto pre-defined SMARTS patterns representing a variety of potential minimal pharmacophores such as two-bond donor-acceptors, acids, lipophiles, etc.

The protocol is as follows: First, for a fragment candidate $A$ predicted to bind to the active site of a protein target, we derive the attribution weight vector $z$ with components $z_a$ (see Methods). Second, we evaluate for each pharmacophore candidate $s$, a vector $m_s$ with components $m_{sa} = \pm 1$ that indicate whether or not an atom $a \in A$ participates in pattern $s$. Third, we rank the different patterns $s$ using our weights $z_A$ according to an AUC $A_s = A(z, m_s)$. In other words, we interpret the attribution weights as scores for classifying individual atomic neighbourhoods into participating ($m_{sa} = +1$) or non-participating ($m_{sa} = -1$) environments. See the Methods section for details.

To visualize the outcome, we project the whitened amplitudes $\hat{A}_s$ onto a two-dimensional map of the minimal pharmacophores. This map is derived from the co-occurrence statistics of these pharmacophores across the fragment library. The “distance” between pharmacophores $s$ and $s'$ is evaluated as $D_{ss'} = 1 - C_{ss'}$, with a correlation $C_{ss'} = \sum_A I_{As} I_{As'} / \sqrt{N_s N_{s'}}$: Here $I_{As} \in \{0, 1\}$ is an indicator function for the presence of pharmacophore $s$ in fragment $A$; $N_s = \sum_A I_{As}$ are total pharmacophore counts.

Fig. 4 exemplifies the resulting pharmacophore heat maps for three systems: (a) the ATP site of cyclin-dependent kinase 2 (CDK2) [23], (b) the protein-protein interaction target KEAP1 [24], (c) the trypsin-like serine protease, urokinase plasminogen activator uPA [25]. As can be seen from the heat maps in the top row, the pharmacophore map displays strong, distinctive hotspots for each system, corresponding to (a) donor-acceptor, (b) acid and acceptor-acceptor, (c) amine and amidine patterns. For comparison, the bottom row of Fig. 4 shows the heat maps derived from the manual pharmacophore annotations: These are clearly more localized and specific than the machine-learned projections. The agreement is adequate, taking into account that some cross-talk between different pharmacophores is to be expected, as the SMARTS definitions are not mutually exclusive: This confusion is particularly visible for the CDK2 kinase in Fig. 4a. Nevertheless, pharmacophore projections of this type may prove useful for...
Conclusions

Even though various studies have investigated the potential of machine-learned models of ligand-protein interactions for drug discovery, attempts to query, assess and visualize the physical interactions learnt and predicted by those models have remained relatively scarce. Here we have validated rigorously and on a large scale the physical binding patterns inferred by state-of-the-art machine learning. Measuring and controlling for both internal consistency and agreement with human experts, our analysis of the attribution outcomes indicates that, overall, the attribution is surprisingly successful in identifying meaningful binding patterns from fragment binding data, even in cases where the raw classification performance is relatively low (AUC < 0.75).

The fact that these models are derived from X-ray hits and misses rather than from binding data on larger more complex molecules improves the machine-learning outcome: The quality of the models benefits greatly from a balanced, diverse fragment library with minimal structural redundancy – design criteria which, if violated, result in models displaying inflated performance measures and an impaired physical “understanding”.

The models may in turn help us improve the composition of fragment libraries through global and class-specific hit-rate modelling. Furthermore, and perhaps most importantly, they provide medicinal chemists and modellers with an unbiased and complete view of the output of a fragment screening campaign, e.g., in the form of hot spots within fragment hits, which can be used to guide subsequent fragment-to-lead optimisation.

Methods

*Data preparation. We consider all crystallographic experiments performed at Astex that involved the soaking or co-crystallisation of a fragment of the Astex X-ray screening fragment library (here referred to as the *Astex fragment library*) that forms a subset of Astex’s larger biophysical screening fragment libraries. The vast majority of these crystallographic experiments would have formed part of a primary X-ray fragment screen. This *Astex fragment library* comprises 1660 validated X-ray hits and 26900 validated X-ray misses against primary and allosteric sites, identified by clustering the hits based on distance- and density-based criteria. This results in an average of 2-3, and a maximum of up to twelve sites per target. For the *XChem dataset*, only the hits have been reported. Therefore, we approximated the *XChem fragment library* as the union of all hits against all targets in the *XChem dataset*. To define the misses for each target, we formed the difference between this *XChem fragment library* and the observed hits for that target. Hence, for the *XChem dataset* we cannot be certain that all the misses are true misses.

**Hit-rate models.** Global and target-class datasets are constructed from the site-specific data. For the global model, we divide the compounds of the fragment library into high vs low-hit-rate compounds. As dividing criterion we require high-hit-rate fragments to have resulted in a hit against at least two targets of a different class, as judged by their EC (enzyme commission) number. This criterion translates into a hit rate of 3% or higher. For the class-specific models, only kinases (EC2.7) were represented among the set of targets with a critical number large enough to warrant modelling.

*Machine learning. The machine-learning framework is based on a local description of atom-centered chemical environments used as input for kernel-based learning and closely related to existing approaches [26, 27]. The model implicitly learns the correlations and co-occurrences of a particular set of atomic environments that set active compounds (hits) apart from inactive compounds (misses). The role of the attribution is then to identify and visualize which correlations and types of atomic environments are key to successful binding. Formulating the predictions in terms of atomic environments greatly simplifies the attribution procedure as is shown further below.

Clearly there is significant freedom in how to choose the atomic description [28], ranging from tensor-based methods [29], graph-convolutional neural networks [30], to hard-coded convolutional descriptors such as atom-centered symmetry functions and SOAP (Smooth Overlap of Atomic Positions [31, 32]). In this work, all we require is that we can construct a smooth but nevertheless sensitive similarity measure from the atomic description. Here we therefore follow a simple approach where the atomic descriptor vector \( \mathbf{x}^a \) is obtained from a basis-set expansion \( \mathbf{x}^a \) of the local neighbourhood of a heavy atom \( a \) in molecule \( A \),

\[
e_{\text{sim}}^a = \sum_{c \in A} w_{c} f(x_{ca}),
\]

where \( w_{c} \) are weights indicating the type \( c \) of atom \( a \) (carbon, hydrogen, etc.), \( f(x) \) is a cutoff function, \( G_{ul}(r_{aa}) \) are radial basis functions of index \( n \) and angular momentum \( l \), and \( Y_{lm}(r_{aa}) \) are real spherical harmonics. The radial basis functions are in turn constructed from Gaussian functions \( g_{n}(r) \) modulated by a frequency-damping function \( \lambda(r) \) that suppresses high-\( l \) components at shorter distances, and a radial decay \( h(r) \) that approaches \( r^{-2} \) for large \( r \) and thus discounts distant neighbours:

\[
G_{ul}(r_{aa}) = g_{n}(r_{aa})h(r_{aa})\lambda(r_{aa}),
\]

\[
\lambda(r_{aa}) = \exp \left( -\frac{2}{\sigma} \frac{r_{aa}}{\pi} \right).
\]

Here \( \sigma = 0.5 \) is an atomic-width parameter that enforces smoothness. Finally, the atomic descriptor is obtained by rotational averaging of the \( Im \)-components:

\[
Z_{\text{av}} = \sum_{c} e_{\text{sim}}^a c_{\text{avclm}}^a.
\]

The above framework, notably the rotational averaging performed at the end, is virtually equivalent to the SOAP formalism [32], but has the advantage that it is faster to evaluate due to the damping functions \( \lambda(r) \) and offers better control over longer-range vs short-range contributions due to the distance decay \( h(r) \). The convolutions leading to \( Z_{\text{av}} \) suggest a “native” kernel \( k_{ab} = \mathbf{x}^a \cdot \mathbf{x}^b / (|\mathbf{x}^a| |\mathbf{x}^b|) \) over atomic environments, which is inherited from SOAP and mimics the rotationally averaged density overlap of two chemical environments.

The molecular representation is constructed implicitly through the choice of kernel function. For intensive properties such as a binding affinity, a best-match procedure [27] (as opposed to, e.g., averaging of the atomic vectors) yields the appropriate molecular kernel:

\[
K_{AB} = \max_p \sum_{a \in A} \sum_{b \in B} k_{ab} p_{ab},
\]

subject to \( p_{ab} = 1/N_B \) and \( \sum_{b} p_{ab} = 1/N_A \).

The constraints with heavy-atom counts \( N_A \) and \( N_B \) of molecules \( A \) and \( B \) enforce competition among the atoms of each fragment for their best-matching partner, and assign to \( p_{ab} \) the role of a permutation matrix. This kernel can be easily decomposed onto atomic contributions if used together with support-vector machines or Gaussian processes. The decision function \( Z_A \) distinguishing between hits (\( z > 0 \)) and misses (\( z < 0 \)) then reads

\[
Z_A = \sum_{B \in T} K_{AB} w_B + Z_0,
\]
with training set $\mathcal{T}$, kernel power $n \geq 1$ and offset $Z_0$.

**Baseline models.** The baseline models use extended-connectivity fingerprints (ECFPs) as implemented by rdkit [33, 34] in combination with kernel support vector machines and logistic regression. As kernel function we choose $k_{AB} = \exp(-|x_A - x_B|/\|w_A\| \|w_B\|)^n$, with some positive power $n \geq 1$. Experience shows that this simple ECFP kernel performs remarkably well in the prediction of molecular properties. The hyperparameters of the baseline models optimized via nested splits are: the bond radius of the ECFP, the kernel exponent $n$, and the regularization strength $C$ of the SVM and logistic regression.

**Attribution and filtering.** The attribution of $Z_A$ onto environments $\{a\}$ is obtained via [26]

$$ z_a = \sum_{B \in \mathcal{T}} \left( K_{AB}^{-1} w_B \sum_{b \in B} k_{ab}@pub \right) + Z_0/\|A\|. \quad (7) $$

One can show (see the SI appendix) that this attribution recipe in fact corresponds to a kernelized version of integrated gradients [16].

There are several approaches to filtering these weights as to improve their robustness with respect to dataset composition. First, for the comparison with the manual annotations, we use a simple rank filter that, for each molecule, performs robustness with respect to dataset composition. First, for the comparison with manual annotations, we use a simple rank filter that, for each molecule, assigns a confidence to each environment $a$ as to whether the sign of $z_a$ is correct: To this end we sample from the training set the distribution $f_-(z)$ for true negatives (aggregating weights from compounds correctly predicted to be a miss) and the distribution $f_+(z)$ for true positives (aggregating weights from compounds correctly predicted to be a hit). Note that these weights are extracted from the training set in a leave-one-out procedure; i.e., the weights contributing to $f_\pm(z)$ are from compounds excluded during the training, under the assumption that the decision function does not change drastically as single hits and misses are left out one by one. Having thus obtained $f_\pm(z)$, we assess negative predicted weights against the cumulative distribution $F_-(z) = \int_z^{\infty} f_-(z')dz'$ and positive weights against $F_+(z) = \int_z^{-\infty} f_+(z')dz'$. For example, if $F_-(z_a)$ is large (approaching unity), this indicates that the positive weight $z_a$ is abnormally large to have occurred in a true negative compound, and should therefore be considered significant. On the other hand, if $F_+(z_a)$ had a value of close to 0.5, then the magnitude of $z_a$ is in no way extraordinary and could have just as well been observed in a miss.

For the pharmacophore reconstruction, as well as the co-localization metrics, we resort to the z-scored weights as the normalized attributions $z_a = (z_a - \mu_a)/\sigma_a$. Here $\mu_a$ and $\sigma_a$ are, respectively, the average and standard deviation of the atomic weights predicted for the training set.

**Co-localization metric.** We define a correlation measure $\varphi(z)$ as the expected magnitude and sign of the weights in the environment of an atomic centre $a$ given that its own attribution weight is $\hat{a} = \tilde{a}_a$. To obtain the random baseline, we construct a null field $\varphi_0$ by randomly permuting the attribution weights among all atomic centres of the point cloud. Examples for the correlation fields $\varphi(z)$ and $\varphi_0(z)$ are included in the SI appendix.

The performance metric for the attribution is subsequently derived as an integral over the positive and negative branches of the correlation field:

$$ \varphi_{\pm} = \pm \hat{a}_{\pm} \frac{1}{2\pi} \int_{-\infty}^{\pm \infty} \varphi(z) \mathrm{d}z. \quad (9) $$

With $x_{\max}$ chosen as two times the standard deviation of the weights $z$ (i.e., $x_{\max} = 2$), the metrics $\varphi_{\pm}$ assume extremal values of $\pm 1$: These are obtained if the weights of the point cloud are "perfectly" correlated in the sense that the background field $\varphi(z)$ seen by all sites $a$ is identical to their own weight $\hat{a}_a$. In practice, however, even a "perfect" model would not be able to achieve $\varphi_{\pm} = \pm 1$, as the boundaries of the hotspots for each fragment hits cannot be expected to align perfectly, and because some hotspots may not be shared by all hits due to different albeit overlapping binding modes.

**Pharmacophore projection.** Due to the small though nevertheless varying size of the fragments and pharmacophores, the SMARTS AUCs are subject to significant statistical fluctuations. We normalize them using the mean $A_0 = 0.5$ and width $\Delta A$ of the distribution of AUCs obtained from randomized instances of the mapping vector $\tilde{m}_s$:

$$ \tilde{A}_s = \frac{\tilde{A}(z, m_s) - A_0}{\Delta A(z, \tilde{m}_s)}. \quad (10) $$

For example, a normalized AUC of $\tilde{A}_s \geq 2$ would indicate strong ("2σ") agreement between the attribution weights and the pharmacophore $s$.

By averaging the whitened AUCs $\tilde{A}_{s,t}$ over all hits $A \in \mathcal{T}$ of a training set of size $N_T$,

$$ \langle \tilde{A}_s \rangle = \frac{1}{N_T} \sum_A \tilde{A}_{s,t}, \quad (11) $$

we obtain a measure for the compatibility between pharmacophore $s$ and the fragment hit observations made for a specific binding site.

**Code availability.** The core library implementing the $G_{\alpha,Y_{10}}$ formalism, kernels and attribution routines is available online at github.com/capecol/gymax. Attribution models have been incorporated into the BenchML model library [35] (see github.com/capecol/benchml). A usage example will be provided in BenchML’s examples folder, located at /benchml/examples/fragml.

**Data availability.** Astex’s fragment library is proprietary and can therefore not be made available. The most recent XChem data on the other hand can be obtained at fragalysis.diamond.ac.uk. A curated subset used in this work can be obtained from github.com/bingqingcheng/linear-regression-benchmarks (see the binding/Xchemfrag subfolder therein).

[1] Reece, D. C., Congreve, M., Murray, C. W. & Carr, R. Fragment-based lead discovery. Nature Reviews Drug Discovery 3, 660–672 (2004). URL http://www.nature.com/articles/nrd1467.

[2] Thomas, S. E. et al. Structure-guided fragment-based drug discovery at the synchrotron: screening binding sites and correlations with hotspot mapping. Philosophical Transactions of the Royal Society A: Mathematics, Physical and Engineering Sciences 377, 20180422 (2019). URL https://royalsocietypublishing.org/doi/10.1098/rsta.2018.0422.

[3] Jhoti, H., Williams, G., Rees, D. C. & Murray, C. W. The 'rule of three' for fragment-based drug discovery: where are we now? Nature Reviews Drug Discovery 12, 644–644 (2013). URL http://www.nature.com/articles/nrd3926-c1.

[4] Vamanathan, J. et al. Applications of machine learning in drug
discovery and development. *Nature Reviews Drug Discovery* **18**, 463–477 (2019). URL http://www.nature.com/articles/s41573-019-0024-5.

[5] Bajorath, J. *et al.* Artificial Intelligence in Drug Discovery: Into the Great Wide Open. *Journal of Medicinal Chemistry* acs.jmedchem.0c01077 (2020). URL https://pubs.acs.org/doi/10.1021/acs.jmedchem.0c01077.

[6] Coley, C. W., Eyke, N. S. & Jensen, K. F. Autonomous Discovery in the Chemical Sciences Part I: Progress. *Angewandte Chemie International Edition* anie.201909987 (2020). URL https://onlinelibrary.wiley.com/doi/abs/10.1002/anie.201909987.

[7] Poelking, C., Amar, Y., Lapkin, A. & Colwell, L. Noisy, sparse, nonlinear: Navigating the Bermuda Triangle of physical inference with deep filtering. *arXiv:1912.04345 [physics, stat]* (2019). URL http://arxiv.org/abs/1912.04345. ArXiv: 1912.04345.

[8] Gómez-Bombarelli, R. *et al.* Automatic Chemical Design Using a Data-Driven Continuous Representation of Molecules. *ACS Central Science* **4**, 268–276 (2018). URLhttps://pubs.acs.org/doi/10.1021/acscentsci.7b00572.

[9] Wallach, I., Dzamba, M. & Heifets, A. AtomNet: A Deep Convolutional Neural Network for Bioactivity Prediction in Structure-based Drug Discovery. *arXiv:1510.02855 [cs, q-bio, stat]* (2015). URL http://arxiv.org/abs/1510.02855. ArXiv: 1510.02855.

[10] Stautelle, B., Brown, N., Von Delft, F. & Deane, C. M. Data Set Augmentation Allows Deep Learning-Based Virtual Screening to Better Generalize to Unseen Target Classes and Highlight Important Binding Interactions. *Journal of Chemical Information and Modeling* **60**, 3722–3730 (2020). URL https://pubs.acs.org/doi/10.1021/acs.jcim.0c00263.

[11] Li, Y. *et al.* Machine Learning Force Field Parameters from Ab Initio Data. *Journal of Chemical Theory and Computation* **13**, 4492–4503 (2017). URL https://pubs.acs.org/doi/10.1021/acs.jctc.7b00521.

[12] Méndez-Lucio, O., Baillif, B., Clevert, D.-A., Rouqué, D. & Wichard, J. De novo generation of hit-like molecules from gene expression signatures using artificial intelligence. *Nature Communications* **11**, 10 (2020). URL http://www.nature.com/articles/s41467-020-18769-w.

[13] Whitehead, T. M., Irwin, B. W. J., Hunt, P., Segall, M. D. & Conduit, G. J. Imputation of assay Bioactivity Data Using Deep Learning. *Journal of Chemical Information and Modeling* **59**, 1197–1204 (2019). URL https://pubs.acs.org/doi/10.1021/acs.jcim.8b00768.

[14] Wallach, I. & Heifets, A. Most Ligand-Based Classification Benchmarks Reward Memorization Rather than Generalization. *Journal of Chemical Information and Modeling* **58**, 916–932 (2018). URL https://pubs.acs.org/doi/10.1021/acs.jcim.7b00493.

[15] McCloskey, K., Taly, A., Monti, F., Brenner, M. P. & Colwell, L. J. Using attribution to decode binding mechanism in neural network models for chemistry. *Proceedings of the National Academy of Sciences* 201820657 (2019). URL https://www.pnas.org/lookup/doi/10.1073/pnas.1820657116.

[16] Sundararajan, M., Taly, A. & Yan, Q. Axiomatic Attribution for Deep Networks. *arXiv:1703.01365 [cs]* (2017). URL http://arxiv.org/abs/1703.01365. ArXiv: 1703.01365.

[17] Sheridan, R. P. Interpretation of QSAR Models by Coloring Atoms According to Changes in Predicted Activity: How Robust Is It? *Journal of Chemical Information and Modeling* **59**, 1324–1337 (2019). URL https://pubs.acs.org/doi/10.1021/acs.jcim.8b00825.
[32] Bartók, A. P., Kondor, R. & Csányi, G. On representing chemical environments. *Physical Review B* **87** (2013). URL https://link.aps.org/doi/10.1103/PhysRevB.87.184115.

[33] Rogers, D. & Hahn, M. Extended-Connectivity Fingerprints. *Journal of Chemical Information and Modeling* **50**, 742–754 (2010). URL https://pubs.acs.org/doi/10.1021/ci100050t.

[34] Landrum, G. https://www.rdkit.org (2006). URL https://www.rdkit.org.

[35] Poelking, C., Faber, F. A. & Cheng, B. BenchML: an extensible pipelining framework for benchmarking representations of materials and molecules at scale (2021). URL https://arxiv.org/abs/2112.02287.
Meaningful machine learning models and machine-learned pharmacophores from fragment screening campaigns

Carl Poelking,1,2,* Gianni Chessari,2 Christopher W. Murray,2 Richard J. Hall,2 Lucy Colwell,1 and Marcel L. Verdonk2

1Department of Chemistry, University of Cambridge, UK
2Astex Pharmaceuticals, Cambridge, UK

SUPPLEMENTARY INFORMATION

Connection to integrated gradients

The attribution weights $z_a$ defined in Eq. 7 of the main text were derived simply and directly from inspecting the decision function $Z_A$. We note here that the same expression can be derived from a kernelized variant of integrated gradients: Sundararajan et al. proposed the following attribution rule for a function $F$ that maps inputs $x$ onto an output $y$ [1]:

$$f_i(x) = (x_i - x'_i) \int_0^1 \frac{1}{\alpha} \frac{\partial F(x' + \alpha(x - x'))}{\partial x_i} d\alpha. \quad (1)$$

Here $f_i(x)$ is the attribution weight assigned to input $x_i$; $x'$ is a baseline input against which the attribution is offset. The factor $1/\alpha$ inside the integral is not part of the original definition, but was added here to ensure the correct attribution of low-order polynomials (noting that for high-dimensional inputs $x_i$ and highly nonlinear functions $F$ the difference is barely noticeable).

In our kernelized approach, it is natural to consider as inputs the pairwise similarities $k_{ab}$ between the atoms $a$ of a probe structure and the atoms $b$ of fragments $B$ in the training set. We thus associate the attribution weight for atom $a$ with

$$z_a = \sum_B \sum_{b \in B} z_{ab}(k) = \sum_B \sum_{b \in B} k_{ab} \int_0^1 \frac{1}{\alpha} \frac{\partial Z_A(\alpha k)}{\partial k_{ab}} d\alpha,$$

where $k$ is the vector of all pairwise kernel values $k_{ab}$. Furthermore accounting for the constant offset $Z_0$ in the decision function (constant terms are ignored by integrated gradients by definition), we thus arrive at the attribution expression of Eq. 7.

Attribution fields

As described in the Methods section, we defined the attribution correlation field $\varphi(\hat{z})$ as a local distance-weighted average over the attribution weights of the atomic centres of the superposition cloud:

$$\varphi(\hat{z}) = \left\langle \frac{1}{Q_a} \sum_{B \neq A} \sum_{b \in B} \hat{z}_a \exp(-\alpha r_{ab}^2) \right\rangle_{\hat{z}_a=\hat{z}}. \quad (3)$$

Here $Q_a = \sum_{B \neq A} \sum_{b \in B} \exp(-\alpha r_{ab}^2)$ is a normalizing factor; $\hat{z}$ as opposed to $z$ are the whitened (i.e., $z$-scored) attribution weights. Intuitively, the field $\varphi(\hat{z})$ measures the expected magnitude and sign of the weights in the neighbourhood of an atomic centre $a$ given that its own attribution weight is $\hat{z} = \hat{z}_a$. Fig. S1 exemplifies $\varphi(\hat{z})$ for nine systems, relative to a random baseline, which corresponds to a null field $\varphi_0$ obtained by randomly permuting the attribution weights among all atomic centres of the point cloud.

Localization of the attribution weights

The degree to which the machine-learned attribution localizes the inferred binding pattern within a fragment hit can unfortunately not be directly controlled, as the locality is the implicit result of the smooth expansion of the atomic environments and atomic kernel, the competitive matching procedure performed on the molecular level, and post-processing of the weights.

For an example and visual guide to how the framework localizes the attribution weights, Fig. S2 shows the superimposed densities and attribution fields for four binding sites taken from the XChem dataset [2]. We discuss briefly the SARS-CoV-2 main protease, for which XChem recently reported the results of a fragment-screening campaign. Fig. S2b shows the superimposed density of the covalently bound ligands on the left together with the machine-learned attribution field on the right: Even though the scaffolds of the fragment hits overlap significantly, the attribution focuses only on the acetyl group that binds covalently to the protein. Naturally this group also turns out to be the one that is spatially best conserved. This information, however, was not made available to the model. The non-covalent hits against the same protease on the other hand present a much more diffuse picture (Fig. S2c): The fragment hits display significant scattering, which can explain why the machine-learned attributions are unable to pinpoint a single hotspot.

Expert annotations

Figs. S3-S5 compare expert annotations, based on minimal-pharmacophore definitions, with machine-learned attributions for a total of eighteen binding sites. Each row in the figures visualizes the outcome for a single site, including, from left
to right: the superimposed ligand density, the manually assigned pharmacophore, the attribution weight density, and the standard deviation of the weight density. The visualizations are projected onto the plane of best fit as derived from the principal-component analysis of the nuclear coordinates of the superimposed structures, where the superposition is derived from aligning the binding site of the protein as resolved by X-ray. The standard deviation is calculated by resampling with replacement (i.e., bootstrapping) the points (i.e., atoms) of the superimposed cloud density. For each sample, the field is re-evaluated on a grid. Finally the component-wise standard deviation is calculated over the set of samples.

[1] Sundararajan, M., Taly, A. & Yan, Q. Axiomatic Attribution for Deep Networks. arXiv:1703.01365 [cs] (2017). URL http://arxiv.org/abs/1703.01365. ArXiv: 1703.01365.
[2] Diamond Light Source. X. C. H. E. M. https://www.diamond.ac.uk/instruments/mx/fragment-screening.html (2020). URL https://www.diamond.ac.uk/instruments/Mx/Fragment-Screening.html.
Figure S1: The attribution correlation fields for nine example systems: The red curves show \( \varphi(\hat{z}) \), with confidence intervals obtained by bootstrapping the set of atoms that form the superimposed ligand density, and re-evaluating \( \varphi(\hat{z}) \) for every bootstrapped sample. The blue curves represent the null background fields calculated for scrambled data, where the atomic attribution weights were permuted randomly among the atoms of the superimposed density.

Figure S2: Superimposed ligand densities and machine-learned attribution fields derived from a subset of the XChem data. The systems shown are (a) Calcium-calmodulin-dependent kinase 1Da, (b) SARS-CoV-2 main protease (covalent hits), (c) SARS-CoV-2 main protease (non-covalent hits), (d) ADP-sugar pyrophosphatase. In each panel, the superimposed density is shown on the left, the mapped attribution field on the right.
Figure S3: (1/3) Comparison between manual and machine-learned annotations of superimposed fragment hits. Each row consists of: left, the superimposed ligand density; centre-left, the manually assigned pharmacophore weights mapped onto this density; centre-right, the attribution field derived from the filtered weights; right, the standard deviation of the predicted attribution field.
Figure S4: (continued, 2/3) Comparison between manual and machine-learned annotations of superimposed fragment hits. Each row consists of: left, the superimposed ligand density; centre-left, the manually assigned pharmacophore weights mapped onto this density; centre-right, the attribution field derived from the filtered weights; right, the standard deviation of the predicted attribution field.
Figure S5: (continued, 3/3) Comparison between manual and machine-learned annotations of superimposed fragment hits. Each row consists of: left, the superimposed ligand density; centre-left, the manually assigned pharmacophore weights mapped onto this density; centre-right, the attribution field derived from the filtered weights; right, the standard deviation of the predicted attribution field.