A Data-Driven Construction of the Periodic Table of the Elements

Michael J. Willatt, Félix Musil, and Michele Ceriotti

1Laboratory of Computational Science and Modeling, IMX, École Polytechnique Fédérale de Lausanne, 1015 Lausanne, Switzerland
2National Center for Computational Design and Discovery of Novel Materials (MARVEL)

Machine-learning of atomic-scale properties amounts to extracting correlations between structure, composition and the quantity that one wants to predict. Representing the input structure in a way that best reflects such correlations makes it possible to improve the accuracy of the model for a given amount of reference data. When using a description of the structures that is transparent and well-principled, optimizing the representation might reveal insights into the chemistry of the data set. Here we show how one can generalize the SOAP kernel to introduce a distance-dependent weight that accounts for the multi-scale nature of the interactions, and a description of correlations between chemical species. We show that this improves substantially the performance of ML models of molecular and materials stability, while making it easier to work with complex, multi-component systems and to extend SOAP to coarse-grained intermolecular potentials. The element correlations that give the best performing model show striking similarities with the conventional periodic table of the elements, providing an inspiring example of how machine learning can rediscover, and generalize, intuitive concepts that constitute the foundations of chemistry.

METHODS

Many machine learning schemes have been used to link structures and properties [15-16], including more or less sophisticated neural networks [17-21]. Based on the few comparative studies that have appeared in the literature [10-22-23], it appears that, when it comes to predicting atomic-scale properties, simple regression techniques such as kernel ridge (Gaussian process) regression perform as well as or better than their more sophisticated counterparts. Given our focus on structure representation and descriptors, in this work we used kernel ridge regression (KRR), in which the properties of a structure \(A\) are written as a linear combination of non-linear kernel functions \(K(\mathcal{X}, \mathcal{X}')\) that evaluate the similarity between two structures, i.e.

\[
y(A) = \sum_M x_M K(\mathcal{A}, \mathcal{X}_M),
\]

where \(\mathcal{X}_M\) correspond to a set of reference atomic structures, and \(x_M\) are weights that can be optimized by minimizing the discrepancy between the predictions \(y(A)\) and the actual values computed on a set of training structures. The details of KRR have been discussed at length elsewhere [24-25] and so here we will focus instead on the definition and optimization of the kernel function.

SOAP in a bra-ket notation.

The representer theorem guarantees that every well-behaved kernel corresponds to a scalar product between feature vectors that associate each input to a point in a – possibly infinite dimensional – Hilbert space, \(K(\mathcal{X}, \mathcal{X}') = \Phi(\mathcal{X})^T \Phi(\mathcal{X}')\) [26]. The Dirac notation provides a convenient formalism to express vectors in a Hilbert space in an
abstract way that is basis-set independent. This makes it very suitable to express results in quantum mechanics, and – due to the similar algebraic structure – in the context of machine-learning based on kernel ridge regression \cite{21,27}, where one can write $K(\mathcal{X}, \mathcal{X}') = \langle \mathcal{X} | \mathcal{X}' \rangle$.

In the SOAP framework \cite{28}, spherical environments centered on each atom in the system are expressed as densities, which are constructed by superimposing Gaussians functions $g(\mathbf{r})$, centered on each atom. We write such an atom density as a position representation of a ket $|X_j\rangle$,

$$\langle \mathbf{r} | X_j \rangle = \sum_i f_c(r_{ij}) g(\mathbf{r} - \mathbf{r}_{ij}) |\alpha_i\rangle,$$

where $\mathbf{r}_{ij}$ is the displacement vector between atoms $i$ and $j$ and we have introduced orthonormal elemental kets $|\alpha\rangle$ that represent the chemical nature of atoms, and a smooth cutoff function $f_c(r)$ that limits the density to a neighborhood of atom $j$. We can collect together the density from all the atoms of the same species to define an element-specific density

$$\langle \mathbf{r} | X_j \rangle = \sum_{\alpha} \psi_{\alpha}^o(\mathbf{r}) |\alpha\rangle.$$

In the original formulation of SOAP \cite{28}, the atom density is expressed by expanding the environmental density in a basis of orthogonal radial basis functions $R_n(r)$ and spherical harmonics $Y_{\mu}^\lambda(\mathbf{r})$,

$$\langle anlm | X_j \rangle = \int d\mathbf{r} R_n(r)Y_{\mu}^\lambda(\mathbf{r}) \langle \mathbf{r} | X_j \rangle.$$

This descriptor is invariant to permutations of atoms of the same kind, and to rigid translations. It is not, however, rotationally invariant, and so the kernel built as the overlap between two environments would not be consistent with one fundamental physical symmetry. To remedy this shortcoming, one can average the kernel over the $SO(3)$ rotation group to obtain

$$K^{(\nu)}(\mathcal{X}_j, \mathcal{X}_k) = \int d\hat{R} \langle \mathcal{X}_j | \hat{R} | \mathcal{X}_k \rangle^{\nu}.$$  

A remarkable result of the SOAP framework is that the descriptors that are associated with this kernel can be computed explicitly. For the case with $\nu = 2$, the SOAP descriptors correspond to the power spectrum,

$$\langle \alpha\alpha' n' l' | \mathcal{X}_j^{(2)} \rangle \propto \frac{1}{\sqrt{2l+1}} \sum_m \langle anlm | \mathcal{X}_j \rangle^* \langle \alpha\alpha'n'l | X_j \rangle,$$

where $^*$ denotes complex conjugation. One can show that

$$\langle \mathcal{X}_j^{(2)} | \mathcal{X}_k^{(2)} \rangle = \sum_{\alpha\alpha' n'l} \langle \mathcal{X}_j^{(2)} | \alpha\alpha' n'l \rangle \langle \alpha\alpha' n'l | X_k^{(2)} \rangle$$

is precisely the rotationally-averaged kernel \cite{28} for $\nu = 2$, and that it captures the 3-body correlations between atoms within the environment \cite{27}.

To complete our summary of the SOAP framework, we should mention that in many applications thus far the $SO(3)$ vectors have been normalised,

$$\left| X_j^{(\nu)} \right| / \sqrt{\langle X_j^{(\nu)} | X_j^{(\nu)} \rangle} \rightarrow \left| X_j^{(\nu)} \right|,$$

and raised to an integer power $\zeta$,

$$\left| X_j^{(\nu)} \right| \otimes \left| X_j^{(\nu)} \right| \otimes \cdots \otimes \left| X_j^{(\nu)} \right| \rightarrow \left| X_j^{(\nu)} \right|,$$

which makes it possible to go beyond the body order implied by $\nu$ in the definition of the environmental kernel, avoiding the complications of higher-order SOAP descriptors.

Having constructed the $SO(3)$ vectors, there are a variety of ways to obtain a global correlation measure between atomic configurations \cite{12}. The simplest approach (which we follow in this work, and is appropriate to learn properties that can be decomposed in atom-centered contributions) is the average kernel

$$K^{(\nu)}(A, B) = \frac{1}{N_A N_B} \sum_{j \in A} \sum_{k \in B} \langle \mathcal{X}_j^{(\nu)} | \mathcal{X}_k^{(\nu)} \rangle,$$

where $N_A$ ($N_B$) is the number of environments in $A$ ($B$).

**GENERALISING THE SOAP ENVIRONMENTAL KERNEL**

The SOAP formalism provides an elegant framework to construct a rotationally-invariant representation of the atomic density that can be used for machine-learning purposes. While the formalism provides a complete representation of structural correlations of a given order within an atomic environment, the quality and the computational cost of the regression scheme can be improved substantially in practice by modifying the representation so that it incorporates some degree of chemical intuition. For instance, the combination of multiple kernels corresponding to different interatomic distances has been shown to improve the quality of ML models\cite{11}, and the use of an alchemical kernel matrix to describe the similarity of different elements has been shown to be beneficial as well \cite{11}.”
Radially-scaled kernels

In a system with relatively uniform atom density, the overlap between environments $\langle \chi_j | \mathcal{X}_k \rangle$ is dominated by the region farthest from the centre. This could be regarded as rather unphysical, since the interactions between atoms decay with distance, and so the closest atoms should give the most significant contribution to the properties, and is reflected in the observation that multi-scale kernels tend to perform best when very low weights are assigned to the long-range kernels \[11, 20\]. Likewise, a scaling of the weights of different atomic distances within an environment has been shown to be beneficial when using ML to predict atomic-scale properties using a different density-based descriptor \[30\].

One could modify SOAP fingerprints to compensate for this effect by multiplying the atomic probability amplitude \[2\] with a radial scaling $u(r)$. For ease of implementation, however, we apply the scaling directly in the definition of $\psi_{\chi_j}(r)$, using $u(r)$ to determine weights associated with each atom in the environment,

$$\langle \alpha \mid \chi_j \rangle = \sum_{i \in \alpha} f_e(r_{ij}) u(r_{ij}) g(r - r_{ij}). \quad (12)$$

While this construction is an accurate realisation of a density scaling only when the width of the atomic Gaussians is small compared to the variation of $u(r)$, it provides a simple way to test the general idea that requires minimal changes to existing SOAP code.\[31\] One should also consider that the atom that sits at the centre of the environment has a special status in the SOAP framework. While atoms in the environment provide information on the structural correlations, the $j$-th atom sits at the centre of the environment by construction. As a consequence, it is best to treat separately the weight $u_0$ associated with the central atom, i.e. to consider

$$\langle r \mid \chi_j \rangle = u_0 g(r) |\alpha_j\rangle + \sum_{i \neq j} f_e(r_{ij}) u(r_{ij}) g(r - r_{ij}) |\alpha_i\rangle. \quad (13)$$

Alchemical kernels

In the presence of multiple elements, the Dirac notation makes it evident that SOAP descriptors consider each element separately, and do not include a notion of similarity between different elements. This makes the computational cost grow steeply with chemical diversity, and makes it impossible to exploit the similar behavior of different elements across the periodic table. In Refs. \[11, 12\] it was shown that extending SOAP with an alchemical kernel $\kappa_{\alpha\alpha'}$ coupling different elements improved the learning efficiency. It however led to a large increase of the computational cost, as it required considering more terms in the scalar product between two descriptors,

$$\langle \chi_j^{(2)} | \chi_k^{(2)} \rangle_{\kappa} = \sum_{\alpha \beta \alpha' \beta'} u_{\alpha \beta} u_{\alpha' \beta'} \langle \chi_j^{(2)} | \alpha \alpha' \beta' \beta \rangle \times \kappa_{\alpha \beta \alpha' \beta'} \langle \chi_k^{(2)} | \beta \beta' \alpha' \alpha \rangle. \quad (14)$$

The bra-ket notation suggests that $\kappa_{\alpha \alpha'}$ serves essentially the purpose of an operator coupling the elements $|\alpha\rangle$ and $|\alpha'\rangle$.

In this spirit, one can write a decomposition of $\kappa$,

$$\kappa_{\alpha \alpha'} = \sum_{J \in 1} u_{\alpha J} u_{\alpha' J}, \quad (15)$$

where the coefficients can be seen as the components of the elemental kets on an “elemental feature” $|J\rangle$, i.e. $u_{\alpha J} = \langle J|\alpha \rangle$. One can then rewrite Eq. \[14\] as

$$\langle \chi_j^{(2)} | \chi_k^{(2)} \rangle_{\kappa} = \sum_{JnJn'l} \langle \chi_j^{(2)} | JnJ'n' \rangle \langle JnJ'n' | \chi_k^{(2)} \rangle, \quad (16)$$

in which we have introduced a partially contracted version of the original fingerprints,

$$\langle JnJ'n' | \chi_j^{(2)} \rangle = \sum_{\alpha \alpha'} u_{J,\alpha} u_{J',\alpha'} \langle \alpha \alpha' | \chi_j^{(2)} \rangle.$$ 

The transformed SO(3) vector components can be written in terms of the components of $|J\rangle$ in the elemental basis, $u_{J,\alpha} = \langle J|\alpha \rangle$. If the number of basis kets $d_J$ is smaller than the number of elements under consideration, then the effective SO(3) vectors occupy a smaller space than $\{\alpha \alpha' \beta' \beta\}$. This low-dimensionality representation of the chemical space can help improve the accuracy of a ML model in the presence of a large number of elements, and can also translate into substantial savings in terms of memory usage and computational effort. It does, however, break the sparsity of the descriptors, which can negatively affect the computational efficiency for some systems.

Note that this transformation can also be expressed directly in terms of the atom density, i.e. one can write

$$\langle Jr | \chi_j \rangle = \sum_{\alpha} u_{J,\alpha} \langle \alpha | \chi_j \rangle. \quad (17)$$

In the case with $d_J = 1$ this formulation is analogous to several recent attempts to mitigate the complexity of ML models including many chemical species \[32, 33\] by representing heterogeneous systems with a single density, and different weights assigned to various elements. Rewriting the SOAP environmental kernel as Eq. \[10\] makes it possible to consider the $u_{J,\alpha} = \langle J|\alpha \rangle$ as optimisable parameters, to improve the performance of the alchemical kernel $\kappa_{\alpha \alpha'} = \sum_{J} u_{J,\alpha} u_{J,\alpha'}$, or to force it to be low-rank. Different strategies can be used to determine the optimal $u_{J,\alpha}$. Here we propose a cross-validation scheme that exploits the scalar-product nature of the SOAP kernel to re-cast one part of the problem as linear regression, \[23\] which we discuss in detail in the Supporting Information.
Multiple-kernel learning

We have shown that by manipulating the form of the SOAP kernel, e.g. by including a radial scaling, by introducing correlations between elements, or by adjusting other hyperparameters, such as the cutoff radius or the shape of the atomic Gaussian functions, it is possible to obtain different perspectives of the structural correlations, and to tune them to give the best possible performance in a regression task. Determining the most effective representation of a given input is typically what deep neural networks are thought to excel at [35], and exploring this possibility will be the subject of future investigation. Remaining in the context of kernel ridge regression, one can attempt a different approach to further improve the performance of the regression. As done in Ref. [11], one can build a composite kernel out of a selection of different models, i.e.,

\[ K(A, B) = \sum_N w_N K_N(A, B). \] (18)

This multiple-kernel model makes it possible to find the best combination of different representations of the atomic environments, using short and long-range, 2 and 3-body, radially-scaled and alchemically-contracted terms. In a Gaussian Process Regression language, each model is meant to contribute \( \sqrt{w_N} \) to the variance of the target property. The weights can be set manually based on an intuitive understanding of how they contribute to a property, or – more simply – optimised by cross-validation. Note that such combined kernels can still be seen as an explicit inner product between descriptors. In other words, taking sums of multiple kernels can be interpreted equivalently as generalisations of kernels, or as generalisations of descriptors that take the form

\[ |X| = \sqrt{w_1} |X^1| \oplus \sqrt{w_2} |X^2| \oplus \ldots, \] (19)

where \( \oplus \) denotes concatenation.

RESULTS AND DISCUSSION

Having discussed different ways SOAP descriptors can be modified to represent in a more efficient way structure-property relations in complex data sets, we now verify what the practical implications of such modifications are. In order put these ideas to the test, we chose two datasets, what the practical implications of such modifications are.

The elpasolite dataset comprises about 11k DFT-optimised quaternary structures with stoichiometry \( \text{ABC}_6 \) elpasolite \( \text{AlNaK}_6 \text{F}_6 \) being the archetype. We have used the elpasolite data set of Faber et al. [40] in which the four elements constituting each structure were chosen from the 39 main group elements H to Bi. The DFT-relaxed geometries of each structure in the elpasolite crystals are almost identical which means that the data set is geometrically uniform but elementally diverse.

Training data selection

For each data set, we randomly selected two subsets: an optimization set (A) to be used to determine the hyperparameters of the model by cross-validation, and the other (B) to be used for training and testing. All of the optimisations discussed in this article (radial scaling, alchemical kernel learning and multiple-kernel learning) were performed on the A set. Once each optimization was performed, we randomly shuffled and partitioned set B multiple times to produce training set and test set pairs. In order to account for the variability of the model accuracy with respect to the composition of the training and test sets, we averaged over the learning curves for each pair to create the figures presented here.

Reduced-dimensionality alchemical kernels

For the elpasolite crystals, our optimization set contained 2k structures and the remainder were used to construct five training and test set pairs at random (6k and 2k structures respectively). Figure 1 shows the averaged learning curves. The reference curve (bright red line) was taken from Ref. [30] and corresponds to recently-proposed density-based descriptors. The dark red, purple and blue curves show the result of optimising the alchemical kernel, which we did by initializing low-dimensional \( u_{j\alpha} \) based on the \( d_j \) principal components of the alchemical kernel,

\[ \kappa_{\alpha\beta} = e^{-(r_\alpha-r_\beta)^2/2\sigma_\alpha^2-(r_\alpha-r_\beta)^2/2\sigma_\beta^2}, \] (20)

where \( r_\alpha \) and \( r_\beta \) correspond to Pauling atomic electronegativity and van der Waals radius for the element \( \alpha \). The values of \( u_{j\alpha} \) were then optimized with an iterative scheme working in the primal formulation of ridge regression for \( \zeta = 1 \) (see SI).
Reducing the dimensionality of the SOAP descriptors by three orders of magnitude with \( d_J = 1 \) leads to a poor learning rate (dark red line). The learning behaviour is much improved with \( d_J = 2 \) (purple line), which corresponds to a reduction in the dimensionality of the SOAP descriptors by a factor of 380. For fewer than 2k structures, the performance is better than standard SOAP (black line), but the learning rate gradually decreases (saturation) as the number of training structures increases. This suggests that the \( d_J = 2 \) descriptor is unable to represent diversity adequately in large sets of structures because of its low dimensionality, in much the same way as reducing \( \zeta \) has been found to lead to saturation in SOAP models trained on the QM9 data set.[31]

By increasing \( d_J \) to 4 (blue line), which corresponds to a reduction in the dimensionality of the SOAP descriptors by 99%, the resulting model outperforms both the reference (bright red line) and standard SOAP models. There is still, however, a reduction in the learning rate as the number of training structures increases. Again, this is likely an indication that the low dimensionality of the descriptor is unable to represent diversity adequately in large sets of structures (in contrast to the higher-dimensional standard SOAP descriptor).

To test this idea, we combined multiple kernels in linear combination, including full-dimensionality standard SOAP kernels for \( r_c = 4, 5, 6 \) and \( \zeta = 1, 2, 3, 4 \), and the optimal alchemical kernels for \( d_J = 1, 2, 4 \). This multiple-kernel model (grey line) combines the optimised elemental correlations of the alchemical descriptor with the resistance to saturation of the standard SOAP descriptor, leading to an improvement in performance over standard SOAP and the state of the art by some 30% on the full training set. It is worth noting that our regression model also outperforms by a factor of two a recently-proposed scheme to determine similarities between elements based on artificial intelligence techniques[42].

Further experimentation with this data set revealed that there is little improvement in model performance when increasing \( d_J \) beyond four (results omitted). Given that a multi-kernel scheme that includes full SOAP models does improve significantly the prediction accuracy, \( d_J = 4 \) does not incorporate all the available chemical information. We attribute the lack of improvements with \( d_J > 4 \) to the failure of our cross-validation scheme to extract meaningful information from the relatively small optimization set.
A data-driven periodic table of the elements

The eigenvectors of the alchemical kernel $K_{\alpha\alpha'}$ lend themselves naturally to be interpreted as spanning a continuous alchemical space in which the elemental kets $|\alpha\rangle$ are embedded. In other terms, they make it possible to obtain a low-dimensional representation of the elements, in which case elements that behave in a similar way with respect to the target property lie close to each other. Figure 2(a) shows the optimized distribution of the elements $u_{\alpha J}$ in the two-dimensional space spanned by $|1\rangle$ and $|2\rangle$ for $d_J = 2$. Elements within different groups of the periodic table are coloured differently. It is immediately apparent from this colouring scheme that optimisation of the alchemical kernel leads to clustering of elements that is reminiscent of their position in the periodic table. The correlation between the data-driven element descriptors and the position in the periodic table is perhaps even more apparent in Fig. 2(b), in which the periodic table is color-coded according to the values of $u_{J\alpha}$. This fascinating observation suggests that one could in principle construct a reasonable alchemical kernel using chemical intuition alone. However, there are two significant advantages to the approach presented here. First, the optimisation is performed automatically on the data set under consideration. Second, the optimisation can be performed just as well in a lower or higher-dimensional space (e.g. $d_J = 1$ or $d_J = 4$, Fig. 2(c) and (d)), where intuition based on the (two-dimensional) periodic table is likely to hinder the performance of the model.

It should also be noted that the elpasolite data set consists of configurations that share the same structure, and span a space that is dominated by elemental correlations, making an optimization that ignores geometric correlations particularly effective. More structurally diverse data sets will imply stronger coupling between geometry and composition, making it advisable to consider more general extensions of the SOAP descriptors to extract comparable insight.

Radial scaling in the QM9 data set

Molecular databases such as the QM9 are less elementally diverse (containing only 5 elements), but contain a broad variety of structures. It has been shown that SOAP kernels can predict with great accuracy the stability of these molecules. However, reaching the best accuracy requires a combination of kernels, as in Eq. 18, with different cutoff radii. The combination of kernels with different length scales has been interpreted in terms of the need for encoding in the kernel the notion of multiple length scales in molecular interactions [31]. The same argument can be applied to the optimization of a radial scaling function $u(r)$ (see Section ), so it should be possible to obtain similar accuracy to a multi-scale kernel by simply optimizing a suitable parameterization of such scaling.

Following Eq. 13, we consider the central atom weight $u_0$ as one hyperparameter, and use a simple functional form with a long-range algebraic decay and smooth behavior at $r \to 0$,

$$u(r) = \frac{2}{1 + (r/r_0)^m}.$$  \hspace{2cm} (21)

We optimized $u_0$, $r_0$ and $m$ using a grid search and 10-fold cross validation over an optimization set of 5,000 randomly-selected molecules. The full set of parameters that we tested is given in the SI. Figure 3 compares the learning curves of conventional SOAP for different cutoff radii with the best radial scaling determined on the A set. Radial scaling leads to a substantial (25%) improvement in the performance of the model. It is important to stress that the results we report here are about 20% better than those in Ref. [11], because we removed the 3,054 structures that failed the SMILES consistency test, as is done by other papers using this data set as benchmark, including Ref. [30]. We also attempted to build a multi-kernel model including both conventional SOAP kernels...
and the best radially-scaled kernels. The improvement we could achieve is marginal, which reinforces the notion that an optimal radial scaling of the descriptor is essentially equivalent to an optimised combination of descriptors with different scales.

Although the QM9 data set exhibits a low degree of composition diversity, one can attempt to further improve the performance of the model by introducing correlations between chemical species. In this case it is necessary to use a $\zeta = 2$ exponent to incorporate many-body interactions in the regression, which makes the application of the primal-based optimization scheme we used for elpasolites impractical. Note that the $u_{\alpha}$ optimized for the $\zeta = 1$ descriptors lead to a degradation of the accuracy when used for the $\zeta = 2$ case. For this reason, and inspired by previous results based on a heuristic determination of $\kappa_{\alpha \beta}$ based on the Pauling electronegativity of the atoms [11], we just used Eq. (20) and performed a grid search to find the optimal values of $\sigma_r$ and $\sigma_c$ (see the SI for more details). Fig. 4 shows that this simple ansatz improves significantly the performance of a SOAP-based KRR model, and also combines with the optimized radial scaling to yield a model which is essentially equivalent in performance to the optimized descriptors of Ref. [43]. The success of the rather primitive form of this feature optimization protocol suggests that a more general strategy in which structural and chemical correlations are tuned simultaneously could improve even further beyond the state of the art.

CONCLUSIONS

Thanks to their mathematically sound, unbiased constructions, SOAP descriptors are particularly well-suited to be extended, incorporating information on correlations between structure, composition and properties. We have given two examples of such extensions, representing the behavior of different chemical species as low-dimensional vectors, and modulating the information content of the descriptors with a radial scaling function. These optimizations improve significantly the performance of SOAP descriptors, matching or surpassing the state of the art on two very different data sets – a chemically diverse set of quaternary solid compounds, and a collection of small organic molecules. The framework we use to simplify the description of atomic species can reduce dramatically the complexity and computational costs of machine-learning models for multi-component systems, and could also be applied to coarse-grained models, in which beads correspond to functional groups, and a reduced-dimensionality description could identify features such as polarity or hydrophobicity.

The exercise of optimizing SOAP descriptors does not only lead to more effective machine learning of molecular and materials stability. As we have demonstrated by re-discovering the periodic table of the elements, and extending it to one and four dimensions, it also makes it possible to extract useful insights from the inspection of the optimal combinations of features. When it comes to the applications of machine learning to chemistry, physics and materials science, accuracy and understanding go hand in hand.

ACKNOWLEDGEMENTS

The Authors would like to thank Gábor Csányi for insightful discussion and comments on an early version of the manuscript. MC and MJW were supported by the European Research Council under the European Union’s Horizon 2020 research and innovation programme (grant agreement no. 677013-HBMAP). FM was supported by the the NCCR MARVEL, funded by the Swiss National Science Foundation.
