



RAPID COMMUNICATION

# Molecular dynamics based antimicrobial activity descriptors for synthetic cationic peptides

MALAY RANJAN BISWAL, SANDHYA RAI and MEHER K PRAKASH\*

Theoretical Science Unit, Jawaharlal Nehru Centre for Advanced Scientific Research, Jakkur, Bengaluru, India  
E-mail: meher@jncasr.ac.in

MS received 9 November 2018; revised 27 December 2018; accepted 31 December 2018; published online 4 February 2019

**Abstract.** There is an urgent need to identify novel antimicrobial drugs in light of the development of resistance by the bacteria for a broad spectrum of antibiotics. Antimicrobial peptides are proving to be an effective remedy to which bacteria have not been able to develop resistance easily. With the goal of progressing towards a rational design of AMPs, we developed a neural network based quantitative model relating their physicochemical properties to their activity. A set of synthetic cationic polypeptides (CAMEL-s) (Mee *et al.* in *J. Peptide Res.* 49:89, 1997) which were studied systematically in experiments was used in the development of our model. Intuitive variables derived from short molecular dynamics simulations in octanol were used as descriptors, resulting in a good prediction of activity and underscoring the possibility of a rational design.

**Keywords.** QSAR; CAMEL-s; molecular dynamics; artificial neural networks; activity prediction; drug design.

## 1. Introduction

Antimicrobial peptides (AMPs) that are part of innate immune system in all multicellular organisms are typically of about 12 to 50 amino acids.<sup>1</sup> These AMPs are typically effective against various pathogens like bacteria, fungi, yeast, viruses and even cancer cells.<sup>2</sup> Conventional antibiotics which target specific bacterial sites sometimes become ineffective by a single mutation at the target site.<sup>3,4</sup> Unlike these antibiotics, AMPs have an entirely different mode of action by attacking the cell membrane of the bacteria.<sup>5</sup> Various mechanistic models have been proposed for the action of AMPs:<sup>6</sup> barrel stave model, toroidal model and carpet model, which result in the formation of pores or breaking up of the membrane into micelles. Bacteria do not develop resistance to these AMPs easily. Therefore there is an urgent need to identify novel AMPs. There is a growing interest in the development of synthetic AMPs that are able to mimic the natural antimicrobial peptides in their functionality.<sup>7,8</sup> But owing to the high cost of

production, poor pharmacokinetics and susceptibility to proteolysis, large scale pharmaceutical application has not been possible yet.<sup>9</sup> Resolving some of these problems as well as developing AMPs against new pathogens requires a rational design approach. Understanding the properties of AMPs would serve as an important tool to design novel drug molecules with the essential features that make them antimicrobial, with possibly a low rate of haemolysis. In this work, we focus on one such quantitative structure-activity relationship (QSAR) models, relating the physicochemical properties of AMPs to their activity.

We performed our analysis on a series of cationic AMPs, also known as CAMEL-s.<sup>10</sup> CAMEL-s were designed by making substitutions in the 15 amino acid long synthetic design. Their activity was experimentally quantified<sup>10</sup> in terms of the minimum inhibitory concentration (MIC) against 24 Gram-positive and Gram-negative bacteria. QSAR methods have been used for modelling activity in various contexts.<sup>11–18</sup> The activity data of CAMEL-s peptide was previously used by other QSAR models.<sup>10,19,20</sup> These earlier models predominantly used either sequence based parameters<sup>21</sup> or in turn depended on other predictive tools such as *in vitro* and *in vivo* aggregation propensities.<sup>22</sup> In our

\*For correspondence

Malay Ranjan Biswal and Sandhya Rai have contributed equally to this work.

work we explore the possibility of predicting activity of CAMEL-s peptides using their structure and dynamics.

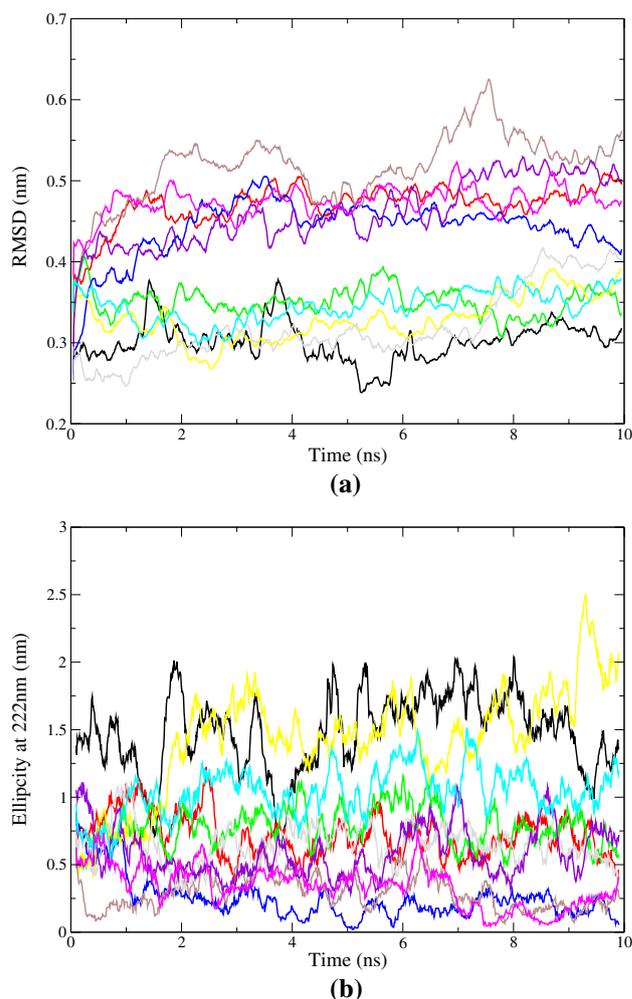
## 2. Methods

### 2.1 Structure and dynamics simulation

The peptides were modelled as  $\alpha$  helical structures using the Modeller tool.<sup>23</sup> Two different sets of molecular dynamics (MD) simulations were performed using GROMACS 5.1.1,<sup>24</sup> one with the peptides solvated in water and another in octanol. For the simulation in octanol, OPLS-AA/L<sup>27</sup> all-atom force field was used for both peptides and octanol. Instead, for the simulation of peptides in water we used AMBER99SB forcefield<sup>25</sup> for protein and TIP3P<sup>26</sup> for water. Most of the average properties converged within 5 ns of MD simulation. RMSD relative to the starting structure and ellipticity at 222 nm calculated using `gmx rms` and `gmx helix` program respectively were used to monitor the convergence. The results shown in Figure 1 show that the systems stabilize after 5 ns. The following parameters were obtained by post-processing the last 5 ns of MD trajectories using the tools available with GROMACS 5.1.1:<sup>24</sup> average dipole moment (`gmx dipoles`), average number of hydrogen bonds within the peptide (`gmx hbond`), solvent accessible surface area (SASA) (`gmx sasa`), ellipticity which is proportional to the number of residues in the helix (`gmx helix`), root mean square deviation, volume (`gmx sasa`), density (`gmx sasa`), moment of inertia along major, middle and minor axes (`gmx principal`). In addition to these, averaged root mean square deviation relative to the starting structure was calculated using VMD.<sup>28</sup>

### 2.2 Neural network model

Artificial neural network based model was set up using scikit learn module<sup>29</sup> in Python 2.7.6. A two layer feed-forward network with logistic sigmoid hidden layer and a linear output layer was used. The network was trained using the low-memory BFGS optimization algorithm. The number of neurons for the hidden layer were varied and the architecture corresponding to the best  $R^2$  value of the validation set was chosen for further analysis. The calculations were repeated using independent calculations with 3 to 5 neurons in the hidden layer. The results we report correspond to the optimal predictions we obtained with 4 hidden neurons. The data was divided into training sets with 81 peptides, validation sets with 10 and test sets with 10 peptides. 50 different training sets were prepared by randomly choosing them from the data. For every pair of training and validation sets, 50 sets of random inputs for initialization of weights and biases of neural network were used, making them a total of 2500 calibration sets. Of these 12 sets were chosen where the  $R^2$  for the training and validation data was greater than 0.6. Based on these,  $R^2$  for the test set was calculated.



**Figure 1.** (a) RMSD of the peptide was calculated with reference to the starting structure and (b) Ellipticity at 222 nm value calculated for some of the peptides solvated in octanol, shows that the physical parameters stabilize after 5 ns. We used data from 5 ns to 10 ns in our analysis.

## 3. Results and Discussion

### 3.1 Dynamical properties

The sequence and biological activity data of 101 CAMEL-s peptides was obtained from Mee *et al.*<sup>10</sup> Starting from the sequence, a putative helical structure was ascribed to each of the peptides, as AMPs mostly assume an  $\alpha$  helical structure, especially in the membrane. The structures were generated using Modeller (Section 2). The dynamics of the peptides in octanol medium, which is believed to mimic the properties in lipid membranes, was investigated using molecular dynamics (MD) simulations and 10 different dynamical properties were calculated as noted in the Section 2. The properties converged with 2 ns as shown in Figure 1 and Figure S1, Supplementary Information. Since biological

activity of proteins or peptides is usually associated with their dynamics, we modelled it using these properties derived from MD simulations. We also repeated the MD with the peptides solvated in water, although as we comment later the predictions were better with octanol simulations.

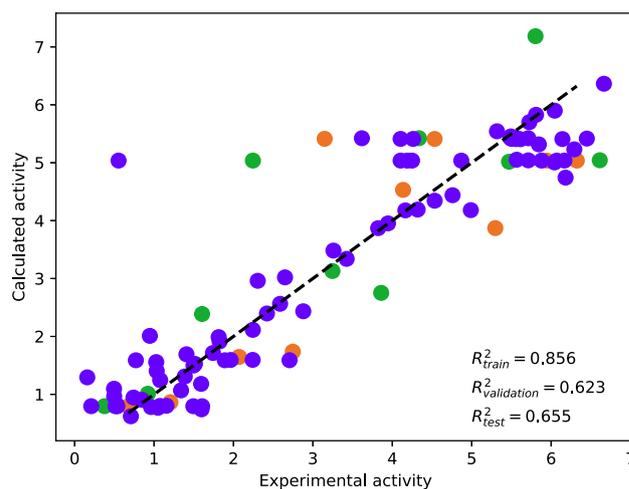
### 3.2 Model for CAMEL-s activity prediction

Ten different variables obtained by processing the MD trajectories (Section 2) were used as descriptors of the activity. These variables included the descriptors of flexibility of the peptide such as RMSD or properties which could describe how easy or difficult an insertion into the membrane might be – such as the average volume or density of the peptide along the trajectory. The activity data was modelled using artificial neural networks (ANN) on these descriptors. The data was randomly split into several sets – 80% training, 10% validation and tested with remaining 10% test set (Section 2). While considering the parameters obtained from simulation in water the neural network model prediction was poor (Figure S2, Supplementary Information). But with simulation in octanol, the model was able to predict the biological activity with fairly good accuracy, the coefficient of determination,  $R^2$  for the training set being 0.86, and that for the validation and test set was 0.62 and 0.65, respectively (Figure 2).

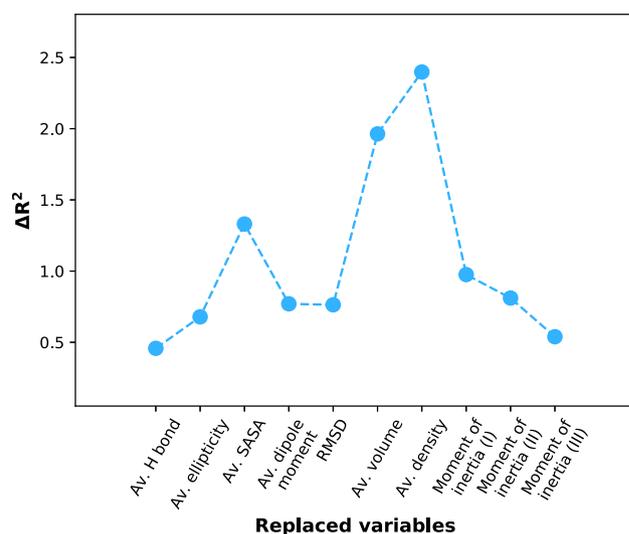
Of the 2500 calibration sets we tried with random initialization (Section 2), 12 resulted in  $R^2$  of training and validation sets to be greater than 0.6. We reported the data set with the best  $R^2$  for the test set. When we repeated the calculations with 70% training, 10% validation and tested with remaining 20%, the results were not as good (data not shown), possibly because of the size of the data set used in the analysis.

### 3.3 Importance of the variables used

In order to understand the relative importance of the ten variables, we compared the  $R^2$  value of the model by replacing variables with their average values. The replaced variables were used as a test data and the  $R^2$  was calculated. The difference between total  $R^2$  of the original and the modified data was calculated and plotted in Figure 3. It is observed that the density of AMPs affects their activity the most, followed by volume and SASA. From Figure 3 it is observed that the density is the most relevant variable and average hydrogen bond is the least. But overall collective effect of the variables helps to determine the activity of the peptides.



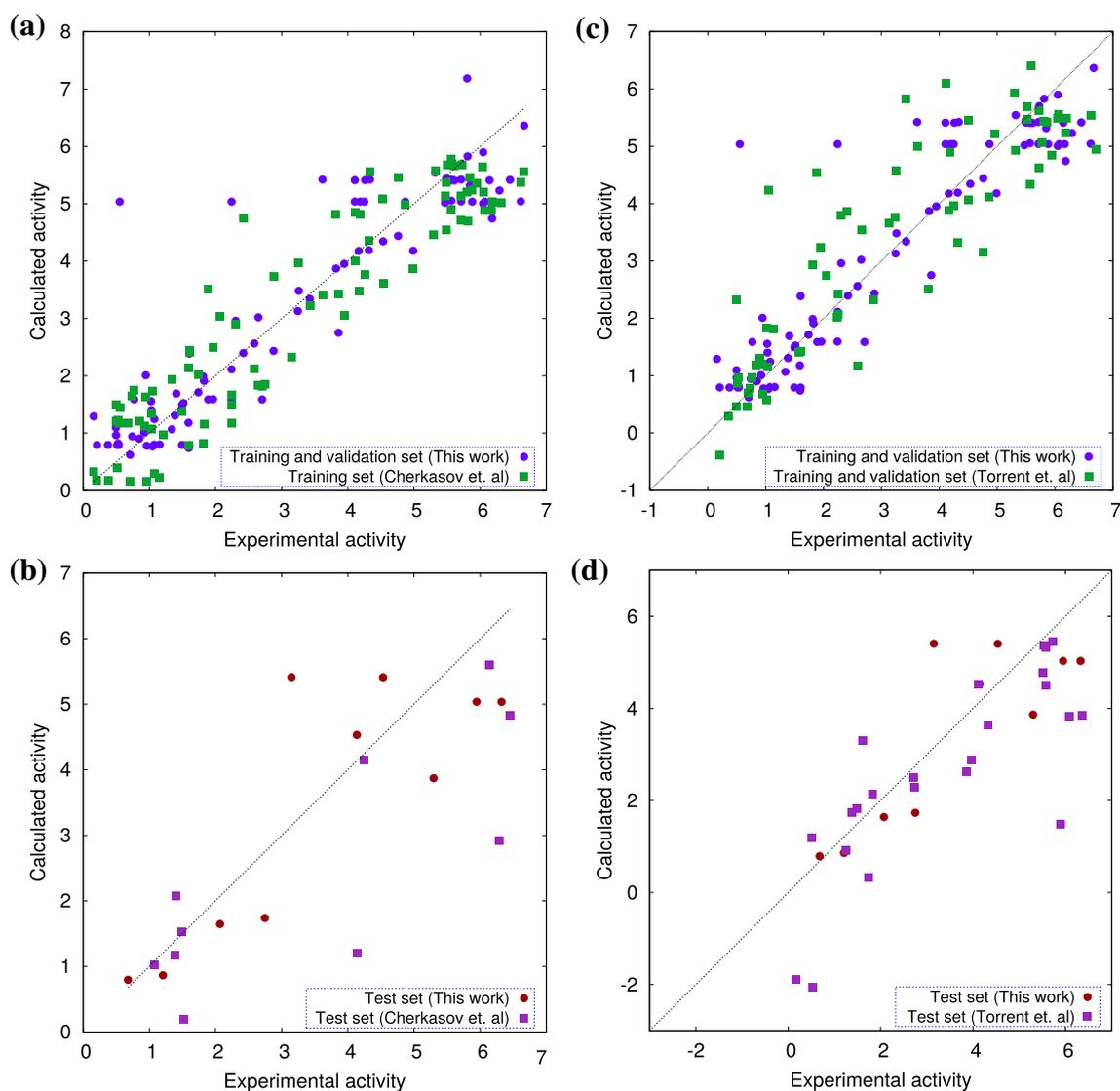
**Figure 2.** The experimental vs. the calculated values of biological activity for the training (purple dots), validation (green dots) and the test set (orange dots).



**Figure 3.** The relative importance of the different variables used in our model was calculated. The difference in  $R^2$  of predictions when each individual variable is replaced by its average value is considered as a measure of importance. Av. denotes average in the figure above.

### 3.4 Discussion

Sequence-structure-dynamics-function is the present paradigm to understand how a sequence translates into a functional peptide or a protein by virtue of its flexible motions in the medium surrounding it. Continuing on this theme, we have used the dynamical properties of AMPs along with artificial neural network models to predict their antibacterial activity. A couple of earlier studies<sup>21,22</sup> attempted the prediction of the activity. We performed the present research with the goal of addressing the following concerns: (i) Since protein/peptide dynamics is believed to be the underlying



**Figure 4.** Comparison between the calculations from our work with those from earlier reported works.<sup>21,22</sup> (a) Calculations from Cherkasov *et al.*, training and validation sets ( $R^2 = 0.86$ ) and (b) test set ( $R^2 = 0.46$ ) (c) Calculations from Torrent *et al.*, training and validation sets ( $R^2 = 0.75$ ) and (d) Test set ( $R^2 = 0.45$ ). Our calculations had  $R^2$  0.86, 0.62 and 0.65 for training, validation and test sets respectively.

factor determining their activity, can the activity be captured using molecular dynamics alone? (ii) Can simple and intuitive variables e.g., volume, density, SASA and RMSD of the peptide be sufficient, as compared to the earlier reported works<sup>21,22</sup> which required derived variables such as aggregation propensities which were in turn predictions from other computational models? (iii) Can a secondary dependence of the descriptive parameters on other predictive tools/software which have not been thoroughly validated be avoided? Our results answer these questions in affirmative, demonstrating the utility of MD for the prediction of AMP activity. The comparison between earlier work and our work is plotted in Figure 4 and it shows that the results of our predictions are at least as good or

slightly better than the predictions from the previous models.

Further, we could see from our calculations that the parameters calculated using MD of peptides in octanol performed better than those from water. The simulations in water resulted in structures which were more flexible than in octanol. The variables density, volume and SASA from the simulations in octanol playing an crucial role in the activity of CAMEL-s peptides and the relatively flexible water models not being representative of the activity suggest the importance of the dynamical state of peptides in octanol, which is usually believed to mimic the membrane environment, are more representative of their functional state.

#### 4. Conclusions

In summary, we could use parameters based on molecular dynamics in octanol to model the activity of antimicrobial peptides on bacteria. The good quantitative results also suggest the utility of peptide dynamics in understanding their biological activity as well as potential for designing novel AMPs.

#### Supplementary Information (SI)

SASA and volume over 10 ns simulation to show the stability of the protein, and prediction using artificial neural network with water are included in the supplementary information. Supplementary Information is available at [www.ias.ac.in/chemsci](http://www.ias.ac.in/chemsci).

#### References

- Jenssen H, Hamill P and Hancock R E W 2006 Peptide antimicrobial agents *Clin. Microbiol. Rev.* **19** 491
- Radek K and Gallo R 2007 Antimicrobial peptides: Natural effectors of the innate immune system *Semin. Immunopathol.* **29** 27
- Fox J L 2013 Antimicrobial peptides stage a comeback *Nat. Biotechnol.* **31** 379
- Narayana J L and Chen J Y 2015 Antimicrobial peptides: Possible anti-infective agents *Peptides* **72** 88
- Brogden K A 2005 Antimicrobial peptides: Pore formers or metabolic inhibitors in bacteria? *Nat. Rev. Microbiol.* **3** 238
- Bahar A A and Ren D 2013 Antimicrobial peptides *Pharmaceuticals* **6** 1543
- Mowery B P, Lee S E, Kissounko D A, Epan R F, Epan R M, Weisblum B, Stahl S S and Gellman S H 2007 Mimicry of antimicrobial host-defense peptides by random copolymers *J. Am. Chem. Soc.* **50** 15474
- Goodman C M, Choi S, Shandler S and DeGrado W F 2007 Foldamers as versatile frameworks for the design and evolution of function *Nat. Chem. Biol.* **3** 252
- Zasloff M 2002 Antimicrobial peptides of multicellular organisms *Nature* **415** 389
- Mee R P, Auton T R and Morgan P J 1997 Design of active analogues of a 15-residue peptide using D-optimal design, QSAR and a combinatorial search algorithm *J. Pept. Res.* **49** 89
- Kubinyi H 2002 From narcosis to hyperspace: The history of QSAR *Quant. Struct. Act. Relat.* **21** 348
- Acharya C, Coop A, Polli J E and MacKerell A D Jr. 2011 Recent advances in ligand-based drug design: Relevance and utility of the conformationally sampled pharmacophore approach *Curr. Comput. Aided Drug Des.* **7** 10
- Favia A D 2011 Theoretical and computational approaches to ligand-based drug discovery *Front. Biosci. Landmark Ed.* **16** 1276
- Mavromoustakos T, Durdagi S, Koukoulitsa C, Simcic M, Papadopoulos M G, Hodoscek M and Grdadolnik S G 2011 Strategies in the rational drug design *Curr. Med. Chem.* **18** 2517
- Verma J, Khedkar V M and Coutinho E C 2010 3D-QSAR in drug design: A review *Curr. Top. Med. Chem.* **10** 95
- Bernard D, Coop A and MacKerell A D 2003 2D conformationally sampled pharmacophore: A ligand-based pharmacophore to differentiate  $\delta$  opioid agonists from antagonists *J. Am. Chem. Soc.* **125** 3101
- Bernard D, Coop A and MacKerell A D 2007 Quantitative conformationally sampled pharmacophore for  $\delta$  opioid ligands: Reevaluation of hydrophobic moieties essential for biological activity *J. Med. Chem.* **50** 1799
- Choudhury C, Priyakumar U D and Sastry G N 2015 Dynamics based pharmacophore models for screening potential inhibitors of mycobacterial cyclopropane synthase *J. Chem. Inf. Model.* **55** 848
- Edlund C, Hedberg M, Engström Å, Flock J I and Wade D 1998 Antianaerobic activity of a cecropin-melittin peptide *Clin. Microbiol. Infect.* **4** 181
- Oh H, Hedberg M, Wade D and Edlund C 2000 Activities of synthetic hybrid peptides against anaerobic bacteria: Aspects of methodology and stability *Antimicrob. Agents Chemother.* **44** 68
- Cherkasov A and Jankovic B 2004 Application of inductive QSAR descriptors for quantification of antibacterial activity of cationic polypeptides *Molecules* **9** 1034
- Torrent M, Andreu D, Nogués V M and Boix E 2011 Connecting peptide physicochemical and antimicrobial properties by a rational prediction model *PLoS One* **6** e16968
- Šali A and Blundell T L 1993 Comparative protein modelling by satisfaction of spatial restraints *J. Mol. Biol.* **234** 779
- Abraham M J, Murtola T, Schulz R, Páll S, Smith J C, Hess B and Lindahl E 2015 GROMACS: High performance molecular simulations through multi-level parallelism from laptops to supercomputers *SoftwareX* **1** 19
- Hornak V, Abel R, Okur A, Strockbine B, Roitberg A and Simmerling C 2006 Comparison of multiple Amber force fields and development of improved protein backbone parameters *Proteins* **65** 712
- Jorgensen W L, Chandrasekhar J, Madura J D, Impey R W and Klein M L 1983 Comparison of simple potential functions for simulating liquid water *J. Chem. Phys.* **79** 926
- Kaminski G A, Friesner R A, Tirado-Rives J and Jorgensen W L 2001 Evaluation and reparametrization of the OPLS-AA force field for proteins via comparison with accurate quantum chemical calculations on peptides *J. Phys. Chem. B* **105** 6474
- Humphrey W, Dalke A and Schulten K 1996 VMD: Visual molecular dynamics *J. Mol. Graph.* **14** 33
- Pedregosa F, Varoquaux G, Gramfort A, Michel V, Thirion B, Grisel O, Blondel M, Prettenhofer P, Weiss R, Dubourg V, Vanderplas J, Passos A, Cournapeau D, Brucher M, Perrot M and Duchesnay E 2011 Scikit-learn: Machine learning in Python *J. Mach. Learn. Res.* **12** 2825