Unraveling the Role of Nanobodies Tetrad on Their Folding and Stability Assisted by Machine and Deep Learning Algorithms

Matheus Vitor Ferreira Ferraz[1,2], Wenny Camilla dos Santos Adan[2], and Roberto Dias Lins[1,2(✉)]

[1] Department of Fundamental Chemistry, Federal University of Pernambuco, Recife, PE 50670-560, Brazil
[2] Department of Virology, Aggeu Magalhães Institute, Oswaldo Cruz Foundation, Recife, PE 50670-420, Brazil
roberto.lins@cpqam.fiocruz.br

Abstract. Nanobodies (Nbs) achieve high solubility and stability due to four conserved residues referred to as the Nb tetrad. While several studies have highlighted the importance of the Nbs tetrad to their stability, a detailed molecular picture of their role has not been provided. In this work, we have used the Rosetta package to engineer synthetic Nbs lacking the Nb tetrad and used the Rosetta Energy Function to assess the structural features of the native and designed Nbs concerning the presence of the Nb tetrad. To develop a classification model, we have benchmarked three different machine learning (ML) and deep learning (DL) algorithms and concluded that more complex models led to better binary classification for our dataset. Our results show that these two classes of Nbs differ significantly in features related to solvation energy and native-like structural properties. Notably, the loss of stability due to the tetrad's absence is chiefly driven by the entropic contribution.

Keywords: Camelid antibodies · Rosetta Energy Function · Machine learning

1 Introduction

Ever since their discovery, single-domain binding fragment of heavy-chain camelid antibodies [1], referred to as nanobodies (Nbs), have gained considerable attention in translational research as therapeutic and diagnostic tools against human diseases and pathogens [2]. Along with its small size (15 kDa) and favorable physical-chemical properties (e.g., thermal and environmental stabilities), Nbs display binding affinities equivalent to conventional antibodies (cAbs) [1, 3]. Moreover, its heterologous expression in bacteria allows overcoming cAbs production pitfalls, such as high production cost and need of animal facility [4, 5]. Hence, Nbs are considered as a promising tool against numerous diseases. A variety of Nbs is currently being investigated under pre-clinical and clinical stages against a wide range of viral infections [6, 7].

J. C. Setubal and W. M. Silva (Eds.): BSB 2020, LNBI 12558, pp. 93–104, 2020.
https://doi.org/10.1007/978-3-030-65775-8_9

The general structural topology of the Nbs is depicted in Fig. 1. It is characterized by a core structure composed of a pair of β-sheets, built from 4 and 5 antiparallel β-strands linked by loops and a disulfide bridge. In contrast to cAbs, which contains six variable loops, Nbs display three highly variable loops H1, H2, and H3. These loops correspond to the Complementary Determining Region (CDR), which is responsible for antigenic binding and recognition, hence providing the target specificity of the Nbs. The overall structure of the Nbs is maintained by four conserved portions, termed as the framework. A significant difference regarding Nbs and cAbs arises from the lack of the variable light chain, and as a consequence, the light-heavy domains interface. To compensate for this loss, four highly conserved residues referred to as the Nb Tetrad are found to replace the nonpolar side chains with polar ones [8, 9]. The Nb tetrad comprises the residues Y/P37, E44, R/C45 and G47. Presumably, these substitutions increase hydrophilicity and solubility of the Nbs, being crucial for their stability [10].

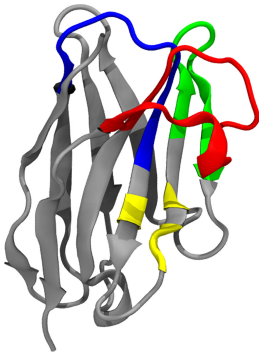

Fig. 1. Cartoon representation of the overall topology of an Nb (PDB ID: 3DWT) [11]. The Nb domain consists of 9 β-sheets linked by loop regions, 3 of these constitute the CDR region and are colored in green, blue, and red. The framework region separated by the hypervariable loops are colored in silver. The Nb tetrad residues are highlighted in yellow. (Color figure online)

These four residues' presence is a hallmark characteristic of Nbs as it has been shown by several sequence alignments studies [12, 13]. The high conservation of these residues indicates an evolutionary-driven constraint, and it highlights their pivotal role in Nbs structure. To ascertain that changes in the Nb tetrad would negatively impact the Nb folding, we have previously designed a Nb by altering the tetrad residues. The obtained chimera presented low expression yields and the absence of a well-defined globular three-dimensional structure due to aggregation (unpublished data). On the contrary, attempts to "camelize" human/murine Abs by grafting the Nb tetrad to the Ab heavy chain's corresponding position has resulted in structural deformations of the framework β-sheet, leading to scarce stability and aggregation [14]. Although it has been described a phage-display library derived from llamas that has produced a set of stable and soluble Nbs devoid of the Nb tetrad [15], these Nbs are unusual, and their stability should be explained in the light of an alternative mechanism.

Given the Nb's tetrad importance in maintaining its folded structure and stability, these residues can be considered key to engineer novel Nbs. It has been shown that molecular dynamics simulations are not sufficient to capture the lower stability of aggregating Nbs, and it does not elucidate the structural and thermodynamic features role of the Nb tetrad to the structures of the Nbs [16]. In this study, we seek to identify the impact of the Nb tetrad from a molecular perspective. To gain insight into the thermodynamic contributions to the folded Nbs, we have used the Rosetta Energy Function components combined with machine learning to identify whether there are differences in the structural pattern of natural Nbs and the corresponding Nbs without the presence of the tetrad sidechains, by replacing them for methyl groups (Alanine). We benchmarked two machine learning (ML) models (Support Vector Machine [17] and Random Forest [18]) and one deep learning (DP) model (Artificial neural network by Multilayer Perceptron [19]) to evaluate the performance of these algorithms in effectively capture the differences among the classes from the multivariate nature of the data.

2 Computational Details

2.1 Dataset Preparation

A total of 30 non-redundant X-ray derived Nb structures, with resolution lower than 3 Å, were retrieved from the Protein Data Bank (PDB). To alleviate bad atomic contacts, the nearest local minimum in the energy function was achieved by geometry-minimizing their initial coordinates using the Rosetta package v. 3.10 [20] and the linear-Broyden-Fletcher-Goldfarb-Shanno minimization flavor conditioned to the Armijo-Goldstein rule. The minimization protocol was carried out in a stepwise fashion, where the sidechain angles were initially geometry-minimized, followed by full rotamer packing and minimization of the orientation of sidechain, backbone, and rigid body. To enhance sampling, $\chi 1$ and $\chi 2$ rotamers angles were used for all residues that pass an extra-chi cutoff of 1. Hydrogen placement was optimized during the protocol. For each of the minimized structures, the Nb tetrad residues were identified and replaced by alanine using the *RosettaScripts* in the four positions, and the obtained structures were geometry-minimized accordingly to the previously described protocol. Thus, the final dataset consisted of 60 instances.

To evaluate folding propensity, the Nb structures were scored using the all-atom Rosetta Energy Function 2015 (REF2015) [21] to calculate the energy of all atomic interactions within the proteins. The REF2015 possesses 20 terms and these were used as the features. The terms can be found in the GitHub (https://github.com/mvfferraz/NanobodiesTetrad), and a detailed description of each term can be found in reference [21]. The score function is a model parametrized to approximate the energy for a given protein conformation. Thus, it consists of a weighted sum of energy terms expressed as mathematical functions based on fundamental physical theories, statistical-mechanical models, and protein structures observations. The Rosetta package is a state-of-art prime-tool to the modeling and design of proteins, and its empirical energy function successfully allows for a valid assessment of the relative thermodynamic stability of folded proteins. The weights for each energy term were kept as default. The parsed command lines, PDB codes, and dataset are available in the GitHub.

2.2 Classification Methods

All the algorithms were written using Python v. 3, and the Scikit Learn Library [22] was employed in conjunction with the Pandas [23] and Numpy [24] packages. In addition, Tensorflow [25] and Keras [26] were used for the NN algorithm. The dataset was split as training (70%) and test (30%) sets. The data features vector was standardized using preprocessing tools to normally distribute the data by scaling the data to a zero mean and unit variance. Details of the code can be found in the GitHub.

Linear Discriminant Analysis (LDA). To identify whether the folding propensity of the Nbs containing the Nb tetrad and those that do not, are linearly separable, a one-component LDA was carried out [27]. LDA projects the input data to a linear subspace constituted of directions to maximize the separation between the classes and minimize the separation among a class. Bayes' statistics are applied to fit conditional class densities for each sample of the data. To select the significant variables, the ensemble learning method of extremely randomized trees (Extra Tree) classifier [28] was used. The number of estimators was kept as 100. The number of features to consider when searching for the best split was assigned as 2, and the quality of a split was measured using the entropy criterion. The LDA was solved using eigenvalue decomposition and was performed by fitting the data and then transforming it without additional parameters. The weights of the LD were used to detect which features are responsible for separating the classes explicitly. To compare if two means were statically different, two-tailed paired t-test was used (GraphPad Prism 8 [29]). Differences were considered statistically significant for a p-value such that $p < 0.05$, at the 95% confidence level.

Support Vector Machine (SVM). SVM consists of a non-probabilistic binary linear classifier, and wherein classification is performed by the construction of a set of hyperplanes in a high-dimensional space. SVM seeks to find a line of separation between the hyperplanes from each class. This line is optimally drawn for maximizing the distance between the closest points regarding each class. C-Support Vector Classification (SVC) was used with a linear Kernel with $C = 1$ hyperparameter, identified with a grid-search over pre-defined values for C (0.001, 0.01, 0.1, 1, 10, 100) and different types of Kernel (Linear and Radial basis function). The linear Kernel, K, is defined as a function of the vectors in the input space, x and y, as $K(x, y) = x^T y$, for $x, y \in \mathbb{R}^d$.

Random Forest (RF). RF is a meta estimator that builds a number of decision trees on bootstrapped training samples and uses averaging from random samples for each split in a tree. All parameters were implemented as the default, save by the criterion to measure the split's quality, set as entropy envisioning information gain.

Neural Network (NN). TensorFlow library was used in conjunction with the Keras high-level application programming interface. The classification was performed using the Multi-layer Perceptron (MLP) Classifier with 100 hidden layers. An MLP is a feed-forward artificial NN class, which learns a function $f(\cdot) : R^m \rightarrow R^n$ by training on a dataset with m input dimensions and n output dimensions, and it contains hidden layers in between the input and output layer. Each hidden layer contains a weight propagated for each posterior layer as a weighted linear summation and followed by a non-linear

activation function $g(\cdot) : R \rightarrow R$. The weight optimization was conducted by stochastic gradient descent, and the step-size for updating the weights was defined as 0.01. Maximum iterations number was set as 500, or until it reaches convergence by considering the default tolerance.

Diagnostic Performance Evaluation. Four performance measures were assessed. The accuracy of the models was computed using the 10-fold cross-validation. To verify the model's performance, the confusion matrix, and the Receiver Operating Characteristic (ROC) curve were evaluated along with the models' learning curve. For a binary classification task, precision, recall, and f1-score are defined according to the assigned classification (true positive (tp), true negative (tn), false positive (fp) and false negative (fn)) as described by Eqs. 1–3. For a detailed description of each metric, see [30].

$$\text{Precision} = \frac{\text{tp}}{\text{tp} + \text{fp}} \tag{1}$$

$$\text{Recall} = \frac{\text{tp}}{\text{tp} + \text{fn}} \tag{2}$$

$$\text{f1} = \frac{2}{\text{recall}^{-1} + \text{precision}^{-1}} = \frac{\text{tp}}{\text{tp} + \frac{1}{2}(\text{fp} + \text{fn})} \tag{3}$$

3 Results and Discussion

3.1 Features Selection

The Rosetta energy terms are convenient mathematical approximations to the physics that governs protein structure and stability. The Rosetta Energy Function (REF) ranks the relative fitness of several amino acid sequences for a given protein structure, and it is capable of predicting the threshold for protein stability by discriminating native-like from non-native structures in a decoy [31]. The functional form relies upon pairwise decomposability of energy terms. The decomposition limits the number of energetic contributions to $1/2N(N-1)$, where N is the atom's number in the system.

When using Rosetta energy function to calculate the score of a protein, *i.e.*, the relative energy for a given conformation reasoned by specific parameters of the Hamiltonian, it yields a total of 20 energetic terms [21]. A feature selection was performed to reduce the effects of noise or irrelevant variables to construct the models. A feature was considered relevant and non-redundant if it presented a feature importance score higher than 0.05 (Fig. 2). From the obtained split, a total of 7 features were filtered:

- *fa_dun*: the probability of a given rotamer is a native-like state based on Dunbrack's statistics for a given ϕ and ψ angles;
- *hbond_sc*: energy for the sidechain-sidechain hydrogen bond;
- *lk_ball_wtd*: asymmetric solvation energy;

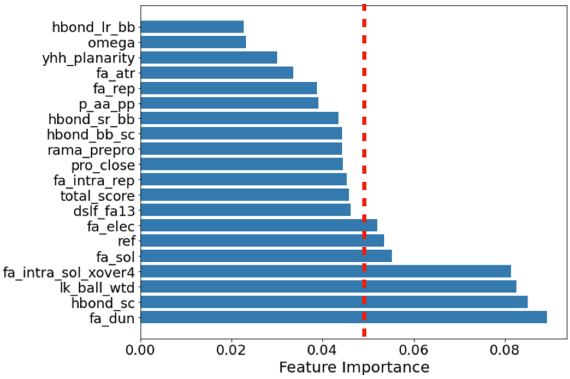

Fig. 2. Feature selection based on REF terms using Extra trees classifier. Feature importance greater than 0.05 was regarded as a relevant feature. The red dashed line represents the threshold for a given feature to be filtered. (Color figure online)

- *fa_intra_sol_xover4*: Lazaridis-Karplus solvation energy for intra-residue interactions;
- *fa_sol*: Lazaridis-Karplus solvation energy model based on Gaussian exclusion;
- *ref*: An approximation to the relative energies of the unfolded-state ensembles;
- *fa_elec*: Coulombic electrostatic potential.

The short descriptions of the terms were retrieved from [21]. As can be seen, almost half of the selected terms are related to the system's solvation properties. Since the replacement of hydrophilic residues for alanine increases the hydrophobic content of the Nbs lacking Nb tetrad, these structural differences have potentially been captured by the REF. These observations corroborate the well-described importance of the Nbs tetrad for solubility.

3.2 Linear Separability of the Data

LDA was used to evaluate whether the filtered features' combination can discriminate natural Nbs from Nbs lacking Nb tetrad. Since we have two classes, the LDA was performed in a one-dimensional fashion. LDA is a supervised dimensionality reductor that identifies the attributes that mostly account for the classes' variance. From fitting a Gaussian density to each class, a single LD was able to separate the class linearly. Figure 3A shows the one-dimensional separability for the classes. In general, natural Nbs lead to a negative value for the LD, and the contrary is observed for the Nbs that lack the Nb tetrad.

To investigate the features that account for the most separation between the classes, the LD loadings were assessed. The loadings indicate the contribution of each feature in predicting class assignment and are shown in Fig. 3B. A higher weight (relatively to their modulus) are *fa_dun*, *lk_ball_wtd*, and *ref*. These results highlight the importance of the Nbs tetrad to solubility and stability of Nbs. The *lk_ball_wtd* consists of the orientation-dependent solvation of polar atoms when assuming the ideal water geometry. As already stated, the REF was able to capture the solvation contribution to the Nb tetrad presence.

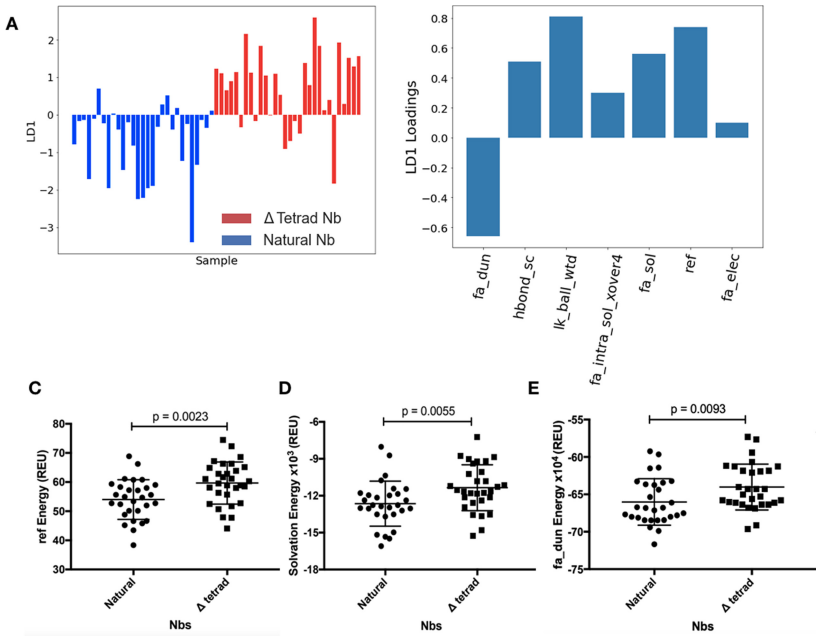

Fig. 3. Discriminant features assessed by LDA. (A) One dimensional LDA. The bars in red are the LD values for the Nbs lacking the Nb tetrad whereas the bars in blue consist of the natural Nbs; (B) Loading of each feature used to calculate de LDs; (C–E) *ref* energy, *lk_ball_wdt*, *fa_dun* energy terms, respectively, for each class. The p-value < 0.05 indicates significant differences in the means. (Color figure online)

Moreover, the other two features are related to native-like conformations properties. Thus, these results show that the Nb tetrad potentially impacts Nbs' backbone ϕ and ψ angles distribution as it is found in the Dunbrack's library of rotamers. Given the importance of the torsion angles for protein folding, a putative explanation for the Nbs tetrad's role in maintaining the structure arises from a geometrical issue. It must be noted that regarding the geometric features, this effect is unlikely to be an artifact from the modeling, since the replacement by alanine residues are not expected to cause significant structural changes, due to the small size of its sidechain, it can be positioned and matched for in any part of the protein (except when replacing tightly buried glycine residues).

To ensure the average of these features were statistically different between the classes, a two-tailed t-test was used. Figure 3(C–E) shows the distribution of the data for each class, along with averages and standard deviations. All three features presented a p-value < 0.05, indicating that these averages are statistically different.

3.3 Classification Algorithms

Given that the selected features can discriminate between the natural Nbs and Nbs that lack the Nb tetrad, classification algorithms were employed to compare the different performance in capturing the classes' structural differences. In this benchmark, machine

learning (SVM and RF) and deep learning (ANN-MLP) were assessed regarding their binary classification performance. SVM is an instance-based learning model, and RF is an ensemble method. MLP is a class of NN, and here it has been employed more than three hidden-layers, and therefore, consists of a DL approach.

All the models have been prepared with the same data and training set. All the 20 features were taken into account to carry out the classifications, since using the selected features from extra trees classifier resulted in poor performances (Data not shown for conciseness). Since it is a small dataset, it is prone to suffer from overfitting the data (high variance). Thus, we performed several performance evaluations. Initially, the models were compared regarding their threshold metrics. Threshold metrics are useful for diagnosing classification prediction errors. Initially, the scores (Fig. 4A), which are directly associated with a combination of the precision and the recall values, were calculated using two approaches: 1) Evaluation was performed considering the initial training/test set; 2) A 10-fold cross-validation was employed. In the latter flavor of evaluating the estimator performance, the training set is split into k sets, and the metrics are calculated in a loop for the different generated sets. The performance is then measured by the average of each k-fold cross-validation. The SVM model presented a remarkable performance in properly assigning the classes, with an accuracy of 0.94 for the initial test set, and an accuracy average of 0.80 when considering ten different subsets. Followed by SVM, MLP also presented good metrics, even though with a slightly lower value. From the three models, the one with the poorest metrics was the RF algorithm. The two formers are more complex and robust models, so that the classification task is likely not trivial, in such a way, a simpler algorithm will not capture the main differences between the classes. The algorithms were compared using the confusion matrix (Fig. 4B–D). The diagonal elements of the matrixes consist of the number of true label classification, whereas, off-diagonal elements represent the mislabeled classifications. The SVM and MLP algorithms outperformed the RF model. The performance metrics are summarized in Table 1 and demonstrate the SVM and MLP algorithms' efficacy for our dataset.

To identify how much the models can benefit from adding more data, learning curves were plotted. Two learning curves were constructed: 1) Train learning curve: calculated based on the training set and diagnosis how well the model is learning, and 2) Validation learning curve: calculated based on a hold-out validation set and diagnosis how well the model is generalizing. Figure 5D–F shows the learning curve for the models. For SVM, the training curve modestly decreases as more samples are added, and the learning curve

Table 1. Threshold performance metrics for each binary classification model

Model	Nb	Precision	Recall	F1-score	Instances	Accuracy
SVM	Natural	0.90	1.00	0.95	9	0.94
	Δ Tetrad	1.00	0.89	0.94	9	
RF	Natural	0.83	0.56	0.67	9	0.72
	Δ Tetrad	0.67	0.89	0.76	9	
MLP	Natural	0.82	1.00	0.90	9	0.89
	Δ Tetrad	1.00	0.78	0.88	9	

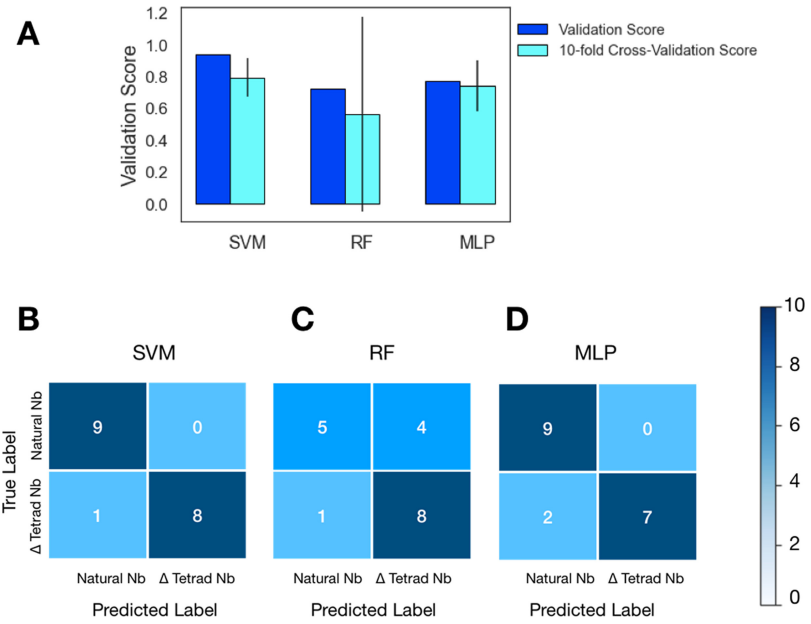

Fig. 4. Threshold metrics for the models' performance (A) Validation scores for the SMV, RF and MLP models considering the accuracy for the train/test split and for the 10-fold cross-validation; (B–D) Confusion matrix for SVM, RF, and MLP

increases until reaching a plateau at a score of nearly 0.80. As can be seen, the model fits the data well, but its generalization has a slightly lower value for the score. Thus, the SVM model might be slightly overfitted. However, its learning capability is progressively increased as more samples are added, indicating that the set number is small. For MLP, a similar trend is observed. However, for the same number of samples, SVM acquires a higher score for the learning curve, suggesting a better model's performance. These results indicate that one source of difficulty for classifying using this dataset resides in the small number of samples.

Furthermore, it shows that the algorithm's training and learning process is not straightforward, given that MLP presents a higher score for the training, proposing that the more complex fitting to the data is required. The RF model did not reflect sensitivity to increasing the number of samples, and a decrease in the learning curve is observed. Thus, the RF model does not benefit from increasing the dataset, and its overfitting cannot be attributed solely to the small size of the dataset, but rather to the simplicity of the algorithm over a complex classificatory task.

These information show that SVM and MLP have the potential to classify between the classes. Such a model is of fundamental relevance for a myriad of protein design algorithms that rely on Monte Carlo sampling. Since a large number of decoys are usually generated, identifying the Nbs that possess native-like characteristics is of enormous advantage to time and resources saving for experimental characterization. From our benchmarking, the RF model is not a proper model to learn from the data. Besides SVM

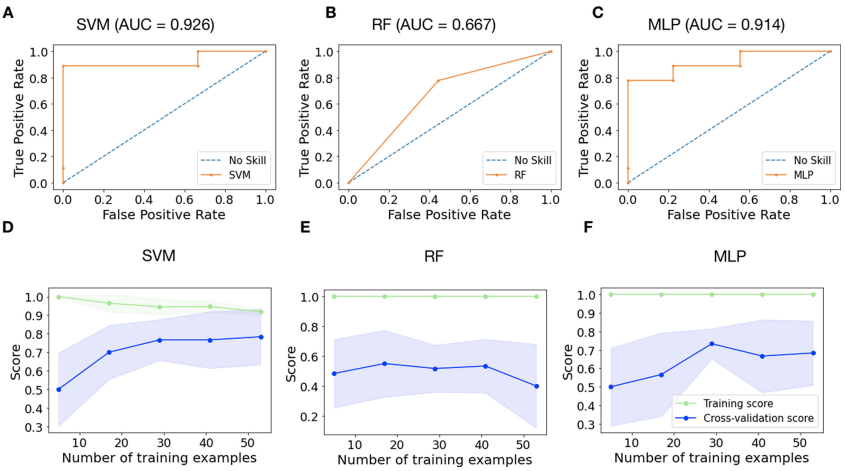

Fig. 5. Assessment of the models' performance through their characteristic curves. (A–C) ROC curves for SVM, RF, and MLP models, respectively; (D–F) Learning curves and the validation score as function of the number of training examples for SVM, RF, and MLP, respectively

having a slight advantage over MLP, the latter is a promising alternative since its training curve perfectly fits the training data, and its increasing learning curve is a promising indicator of its potential. The SVM presented a satisfactory performance, and from searching for different parameters combination, a considerable gain in the predictivity capacity might be observed. The ML and DL algorithms' performance confirms that there are traits that allow for the discrimination of Nbs containing the tetrad or not. Our results show that abolishing the tetrad associates with loss of folding stability in agreement with literature data. It is captured by the *ref* term, which in turn is shown to have significant contributions from the solvation energy and torsional dihedral motion terms. Therefore, the loss of Nbs stability due to eliminating the tetrad is a mostly entropic-driven phenomenon.

4 Conclusions

We have compared the structural features, calculated by the REF's energy term, of natural Nbs containing the Nbs tetrad and a synthetic set of Nbs lacking the tetrad. Data mining analyses revealed that the two classes of nanobodies differ mainly by folding and solvation features, corroborating with previous studies suggesting the tetrad's importance for stability and solubility. This work's findings expand the knowledge on the impact of the Nbs tetrad from a molecular-level perspective by highlighting the importance of entropic contributions to their stability.

Acknowledgements. This work has been funded by FACEPE, CAPES, CNPq, and FIOCRUZ. We acknowledge the LNCC for the availability of resources and support.

References

1. Muyldermans, S.: Nanobodies: natural single-domain antibodies. Annu. Rev. Biochem. **82**, 775–797 (2013)
2. Mir, M.A., Mehraj, U., Sheikh, B.A., Hamdani, S.S.: Nanobodies: The "Magic Bullets" in therapeutics, drug delivery and diagnostics. Hum. Antib. **28**, 29–51 (2020)
3. Vincke, C., Muyldermans, S.: Introduction to heavy chain antibodies and derived Nanobodies. Methods Mol. Biol. **911**, 15–26 (2012)
4. Morrison, C.: Nanobody approval gives domain antibodies a boost. Nat. Rev. Drug. Discov. **18**, 485–487 (2019)
5. Jovčevska, I., Muyldermans, S.: The Therapeutic potential of Nanobodies. BioDrugs **34**(1), 11–26 (2019). https://doi.org/10.1007/s40259-019-00392-z
6. Beghein, E., Gettemans, J.: Nanobody technology: A versatile toolkit for microscopic imaging, Protein–Protein interaction analysis, and protein function exploration. Front. Immunol. **8**, 771 (2017)
7. Konwarh, R.: Nanobodies: Prospects of expanding the Gamut of neutralizing antibodies against the novel coronavirus, SARS-CoV-2. Front. Immunol. **11**, 1531 (2020)
8. Revets, H., De Baetselier, P., Muyldermans, S.: Nanobodies as novel agents for cancer therapy. Expert. Opin. Biol. Ther. **5**, 111–124 (2005)
9. Muyldermans, S.: Single domain camel antibodies: Current status. J. Biotechnol. **74**, 277–302 (2001)
10. Barthelemy, P.A., et al.: Comprehensive analysis of the factors contributing to the stability and solubility of autonomous human VH domains. J. Biol. Chem. **283**, 3639–3654 (2008)
11. Vincke, C., Loris, R., Saerens, D., Martinez-Rodriguez, S., Muyldermans, S., Conrath, K.: General strategy to humanize a camelid single-domain antibody and identification of a universal humanized nanobody scaffold. J. Biol. Chem. **284**, 3273–3284 (2009)
12. Mitchell, L.S., Colwell, L.J.: Comparative analysis of nanobody sequence and structure data. Proteins **86**, 697–706 (2018)
13. Kunz, P., et al.: Exploiting sequence and stability information for directing nanobody stability engineering. Biochim. Biophys. Acta Gen. Subj. **1861**, 2196–2205 (2017)
14. Rouet, R., Dudgeon, K., Christie, M., Langley, D., Christ, D.: Fully human VH single domains that rival the stability and cleft recognition of camelid antibodies. J. Biol. Chem. **290**, 11905–11917 (2015)
15. Tanha, J., Dubuc, G., Hirama, T., Narang, S.A., MacKenzie, C.R.: Selection by phage display of llama conventional V(H) fragments with heavy chain antibody V(H)H properties. J. Immunol. Methods **263**, 97–109 (2002)
16. Soler, M.A., de Marco, A., Fortuna, S.: Molecular dynamics simulations and docking enable to explore the biophysical factors controlling the yields of engineered nanobodies. Sci. Rep. **6**, 34869 (2016)
17. Hearst, M.A., Dumais, S.T., Osuna, E., Platt, J., Scholkopf, B.: Support vector machines. IEEE Intell. Syst. Appl. **13**, 18–28 (1998)
18. Breiman, L.: Random forests. Mach. Learn. **45**, 5–32 (2001)
19. Pal, S.K., Mitra, S.: Multilayer perceptron, fuzzy sets, classifiaction. IEEE. Trans. Newural. Netw. **3**(5), 683–697 (1992)
20. Leaver-Fay, A., et al.: ROSETTA3: An object-oriented software suite for the simulation and design of macromolecules. Methods Enzymol. **487**, 545–574 (2011)
21. Alford, R.F., et al.: The Rosetta all-atom energy function for macromolecular modeling and design. J. Chem. Theory Comput. **13**, 3031–3048 (2017)
22. Pedregosa, F., et al.: Scikit-learn: Machine learning in Python. J. Mach. Learn. Res. **12**, 2825–2830 (2011)

23. McKinney, W.: Data structures for statistical computing in python. In: Proceedings of the 9th Python in Science Conference, pp. 56–61. Austin (2010)
24. Harris, C.R., et al.: Array programming with NumPy. Nature **585**, 357–362 (2020)
25. Abadi, M., et al.: Tensorflow: Large-scale machine learning on heterogeneous distributed systems. arXiv preprint arXiv:1603.04467 (2016)
26. Gulli, A., Pal, S.: Deep learning with Keras. Packt Publishing Ltd, Birmingham (2017)
27. Fisher, R.A.: The use of multiple measurements in taxonomic problems. Annals Eugen. **7**, 179–188 (1936)
28. Geurts, P., Ernst, D., Wehenkel, L.: Extremely randomized trees. Mach. Learn. **63**, 3–42 (2006)
29. Prism, G.: Graphpad software. San Diego, CA, USA (1994)
30. Powers, D.M.: Evaluation: From precision, recall and F-measure to ROC, informedness, markedness and correlation. J. Mach. Learn. Technol. **2**, 37–63 (2011)
31. Cunha, K.C., Rusu, V.H., Viana, I.F., Marques, E.T., Dhalia, R., Lins, R.D.: Assessing protein conformational sampling and structural stability via de novo design and molecular dynamics simulations. Biopolymers **103**, 351–361 (2015)