
**A Review of Secondary Structure Neural Networks and Self-Organizing Map Spectroscopy
for Protein Secondary-Structure Prediction**

*Adewale Olamoyesan, Charles Okechukwu Ugwunna, Ikechukwu Ugbaga Nkole,

Ikenna Bennedict Onyeachu

College of Science and Computing, Wigwe University, Isiokpo, Rivers State, Nigeria

*Corresponding Author: adewale.olamoyesan@wigweuniversity.edu.ng;

adewale.olamoyesan@gmail.com

O

ABSTRACT

Self-organizing maps (SOMs) are unsupervised machine-learning methods that project high-dimensional data onto a lower-dimensional representation while preserving the underlying relationships within the data. This approach was applied to protein structure prediction and other practical problems in structural biology. In this minireview, a historical overview was provided of the development and application of SOM-based approaches to protein secondary-structure prediction, followed by a discussion of the requirements for developing and implementing Self-Organizing Map Spectroscopy (SOMSpec). The scientific and practical problems that SOMSpec addressed over more than a decade since the development of its first application were examined, highlighting its contributions to protein structural analysis. The current challenges and limitations associated with SOMSpec were further discussed, including considerations related to reference datasets, spectral quality, model development, and prediction accuracy, and potential directions for its future development were outlined. By providing this historical and methodological perspective, this article aimed to promote wider awareness and adoption of SOM-based spectroscopic approaches for protein structure prediction and to encourage further investigation of SOMSpec by researchers and the biopharmaceutical industry.

Keywords: Self-organising map; Self-organising map spectroscopy; Secondary structure neural network.

INTRODUCTION

Protein secondary structure is a fundamental determinant of protein folding, and biological function [1,2]. Accurate characterization of secondary structural elements such as α -helices, β -sheets, turns, and disordered regions is therefore essential in structural biology, biophysics, and pharmaceutical

development. Although high-resolution techniques such as X-ray crystallography, nuclear magnetic resonance (NMR) spectroscopy, and cryo-electron microscopy provide detailed structural information, they are often time-consuming, expensive, and unsuitable for routine and high-throughput analyses [3-6]. Consequently, spectroscopic methods, particularly circular dichroism (CD), infrared (IR), and Raman spectroscopy, have become important tools for rapid assessment of protein conformation and structural changes [6-9].

The usefulness of these spectroscopic approaches depends heavily on computational methods capable of converting spectral data into quantitative structural information. Traditional methods such as CONTIN, SELCON, CDSSTR, and Bayesian approaches rely on statistical fitting of unknown spectra against reference proteins with established structures [10-14]. While these methods are widely used, their performance may depend on factors such as reference set composition, wavelength range, and assumptions embedded within the fitting process [13,14]. Machine-learning approaches have emerged as attractive alternatives because they can identify complex relationships within multidimensional datasets. Among these, the Self-Organizing Map (SOM), introduced by Kohonen in 1982, offers an unsupervised learning framework that projects high-dimensional data into a lower-dimensional representation while preserving topological relationships among input vectors [15]. This capability makes SOM particularly suitable for protein spectroscopic analysis, where subtle spectral differences may reflect significant structural variations.

The application of SOM methodology to secondary-structure prediction has evolved considerably over the past three decades. Beginning with K2D and SOMCD, these algorithms were subsequently developed into the Secondary Structure Neural Network (SSNN) and later generalized into SOMSpec [16-20]. Unlike earlier implementations, SOMSpec accommodates multiple spectroscopic techniques, including CD, IR, and Raman spectroscopy [20-23].

This paper reviews the historical development of SOMs for secondary structure prediction, describes what SOMs are and the requirements for their implementation, highlights the practical problem-solving capabilities of our SOM models, and discusses the future prospects of SOMSpec. This text would enable researchers and the biopharmaceutical industry to readily implement SOMSpec with knowledge of how reliable it may be for different types of proteins.

EVOLUTION OF SOM-BASED METHODS FOR PROTEIN SECONDARY STRUCTURE PREDICTION

The first significant implementation of SOM methodology for protein secondary structure was the K2D algorithm developed by Andrade and co-workers [16]. K2D employed SOMs to estimate protein secondary structure from CD spectra recorded between 200 and 240 nm. While the method demonstrated the potential of machine learning for spectroscopic structure determination, it was restricted by its limited wavelength range and by its focus on only the major secondary-structure classes and random coil. To improve predictive performance, particularly for β -sheet-rich proteins, K2D was later extended into K2D3 through the incorporation of a larger database of theoretically derived CD spectra [17].

Further refinement of SOM-based approaches led to the development of SOMCD, which expanded both the spectral range and the structural categories considered during prediction [18]. Compared with earlier implementations, SOMCD utilized improved reference datasets and included turn structures in addition to α -helix, β -sheet, and unordered conformations. These developments demonstrated that SOM-based methods could provide reliable structural estimates while maintaining the visualization advantages inherent to self-organizing maps.

A major advance was achieved with the introduction of the Secondary Structure Neural Network which adopted SOM methodology but offered greater flexibility regarding wavelength selection and reference-set composition [19,24,25]. The next stage in this evolutionary pathway was the development of SOMSpec, which generalized the SSNN framework beyond circular dichroism spectroscopy [20]. Unlike SSNN, SOMSpec accepts data from multiple spectroscopic techniques, including CD, infrared (IR), and Raman spectroscopy, thereby extending the utility of SOM-based structural analysis. Additional features, such as leave-one-out cross-validation and visualization tools for map interrogation enhanced its applicability as a general-purpose platform for protein secondary-structure prediction.

Collectively, the progression from K2D to SOMCD, SSNN, and ultimately SOMSpec illustrates the continuous evolution of SOM-based methodologies from specialized CD-analysis tools into versatile machine-learning frameworks capable of extracting structural information from multiple spectroscopic techniques. These developments have expanded the scope of protein

secondary-structure determination and established SOM-based approaches as valuable complements to conventional statistical and spectroscopic analysis methods.

SOM AND THE REQUIREMENTS FOR SETTING UP SSNN OR SOMSPEC

Unlike classical neural networks, the SOM learns to classify high-dimensional data into a low-dimensional representation without supervision, as implied by its name. This algorithm is also known as a Kohonen map, named after its developer, Finnish professor Teuvo Kohonen. SOMs typically do not have hidden layers, and in most applications their output layer consists of a rectangular two-dimensional (2D) grid onto which the transformed data are projected [15]. Alternatively, hexagonal grids are used to provide even distances between the nodes [15]. Handful of authors have implemented SOMs in one dimensional, 1-D (linear) [26] and three-dimensional, 3-D (toroid SOM for example is used to mitigate the effect of inconsistent neighbourhood set, another example of this form is cylinder topology [27, 28]).

SOM generally works by accepting input vectors and auto-organised them into a clustered map. This map acts as visual aid to user(s) as it illustrates the topographical relationships between elements of the input vector(s). Each input vector assigned to a node in the SOM is associated with a corresponding weight vector that shares the same features [29], as expressed in Equation 1:

$$W = (w_{i1} + w_{i2}; \dots, w_{im}), \text{ where } i \text{ or } j = 1, 2, \dots, m \quad (1)$$

where m represents the total number of nodes in the map, while i and j are unit vectors along the x- and y-axes, respectively, in the vector space (i.e., the (x,y) plane). After initialisation of the weighing vectors, all nodes in the neighbourhood of the target property value (input value) compete for the best vectors (most known as best matching unit, BMU) and auto-organised themselves into a clustered map [29]. In general, the BMU at each iteration is determined by using discriminant function that measures the squared Euclidean distance between the input vector $x(t)$ and the weight vector $w(t)$ as shown in Equation 2:

$$d_{ij} = \sqrt{\sum_{ij=1}^m (x_{ij}(t) - w'_{ij}(t))^2} \quad (2)$$

where m is the same as previously defined. Other measures of distance and similarly used in SOMs and input patterns that are well liked are: correlation; direction cosine; block distance [30]. This

step is regarded in most texts as competitive step where learning is based on similarity measuring or matching. Next step involves updating the neighbourhood nodes within or toward the edges of the BMU with similar coordinates (x, y) in the topology of the SOM. The neighbourhood nodes similarity with the BMU is further increased by updating the weight vectors by using the expression (Equation 3):

$$W_i(t + 1) = w_i(t) + \theta(t)L(t)[I(t) - w_i(t)] \quad (3)$$

where t denotes interaction/current time step, $\theta(t)$ the radius bias function, $L(t)$ the learning rate (it dictates the speed at which the weights are altered and convergence at 0), $w_i(t)$ the BMU. A special factor responsible for bringing the neighbourhood nodes closer to the BMU is the learning rate (L), in SSNN or SOMSpec the type of learning rate used is exponential series expressed in Equation 4:

$$L(t) = L_0 e^{-k_1/t} \quad (4)$$

where L_0 represents the initial learning rate, t the current iteration, and k_1 a parameter of size such that it is $\ll 1$. This learning rate has been demonstrated to decrease exponentially with iteration (convergence to 0 or 1 guarantees stability of the SOM), and other learning functions commonly used are linear, and Gaussian [30,31]. Neighbourhood radius changes at each iteration around the winner node is modeled with the expression in Equation 5 in SSNN or SOMSpec [32]:

$$r(t) = \begin{cases} (r_0 - 1) \cdot \left(1 - \frac{t}{t_1}\right) + 1 & \text{if } t \leq t_1 \\ 1 & \text{if } t > t_1 \end{cases} \quad (5)$$

where r represents the neighborhood radius, r_0 the initial radius which is defined as half the size of the map, and t_1 a parameter that indicates when the radius reaches 1, at this point the radius remain the same as the iteration continue.

The final step assigns structural labels to best vector, its neighbours, and other nodes, which are displayed on the output map. For each spectrum in the reference set, the structural labels add up to 1, and this must also be true for the predicted spectrum. After training, the generated map is used to test unknown spectra, with the number of BMUs set to 5. In the test spectrum, nodes similar to the BMUs are determined using a model that quantifies the deviation between experimental and

expected values, defined as the normalized root mean square deviation (NRMSD) (Equation 6). Interpolation of the unknown input across the BMUs generated a predicted spectrum, which is produced as an output alongside the map [32].

$$NRMSD = \frac{\left(\sqrt{\frac{\sum_i (x_{i,experiment} - x_{i,model})^2}{N}} \right)}{(M-m)} \quad (6)$$

where x_i denotes the intensity at each wavelength, N the total number of data points, and M-m the difference between the is the maximum and the minimum intensity values.

As a rule of thumb, based on prior experience with SOMs in protein secondary structure determination and other applications, thousands of iterations are required to complete training of a map. Validation studies of SSNN and SOMSpec for protein secondary structure determination from CD and IR spectra [19,20,24,25] indicate an optimal iteration range of 20,000–50,000, subject to other parameters affecting prediction quality and reliability. For example, SOMSpec validation across map sizes of 10–100 (100–10,000 nodes) and 10,000–100,000 iterations showed NRMSD variation below 10%, with prediction outcomes remaining highly consistent (~1%) upon repeated runs on the same dataset [20].

The procedural steps that take place during the use of SOM for structure and spectral fitting of unknown proteins CD spectra or when applied for other purposes are illustrated by the flow diagram in Fig. 1. User(s) first (i) initiates the program or application designed to create SOM; (ii) enters training parameters: map size(N) to create a grid with N*N nodes, number of iterations to repeat the steps or processes until its convergences after certain number of times, BMU; (iii) initializes the first model in order for the algorithm to run next; (iv) randomly choose an input vector out of the input vectors; (v) computes how similar or dissimilar the input vector is from the model using discriminating factor (Euclidean distance) to find the BMU; (vi) BMU is determine from the nodes similar to the input vector; (vii) determines nodes that are quite similar with the BMU (neighbourhood nodes), if there are none the process return to previous step, if otherwise it proceed to the next step; (viii) calculates the neighbourhood function of BMU and the rate at which it learn; (ix) updates the BMU and its neighborhood nodes to more closely match the input vector;

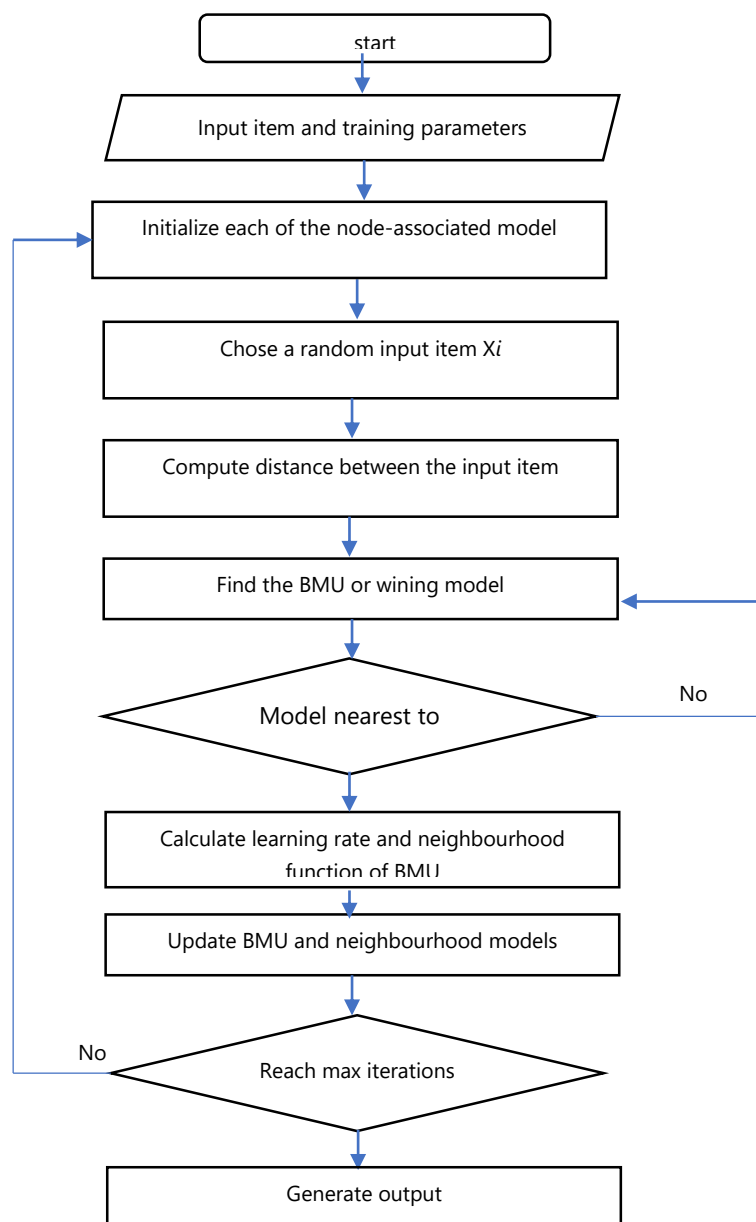


Fig. 1. Typical flow diagram highlighting the procedural steps for creating SOM which can be used to predict the properties of interest from the CD/IR spectra and other data set.

; (x) repeats all the steps for all input items ($x_i - x_n$) until it reaches the maximum iteration or map converges [33].

PROBLEMS SSNN AND SOMSPEC HAVE PROVIDED ANSWERS TO SSNN

The next phase following the development of SSNN involves its implementation for the prediction of protein secondary structure. SSNN was first applied to assess prediction reliability using leave validation approach (leave-one-out cross-validation, LOOCV) on 48 reference set proteins with known structures, *e.g.*, lysozyme, hemoglobin, insulin, and myoglobin. A training map was generated using 47 of the 48 reference proteins and tested on the excluded protein without including its structural information. This process was repeated such that each protein in the reference set served once as the test file. Each training map was created by adopting the optimised parameter values (map size = 40x40, number of iterations = 28000, initial neighbourhood = 20, BMUs = 5, $L_o = 0.1$, $t_1 = 7,000$ iterations, and a $k_1 = 5 \times 10^{-6}$). The structural estimates from SSNN were compared with those from SELCON3, K2D and SOMCD all ran in LOOCV manner. Data obtained from this approach indicated that SSNN predictions were 25 times out of 48 more reliable than that from SELCON3. It appears that the overall performance is almost the same as SELCON3 but significantly better than K2D and SOMCD [25]. Moreover, SSNN was applied to real data to estimate the true concentration of sample protein in a solution and its corresponding secondary structure. There is low confidence in the concentration of the proteins due to evaporation and presence of other molecule(s) that interferes with protein assay and affects its amount in the solution.

The reference set used for creating previous SOM was extended with 5 protein spectra representing 100% α helical and 100% random coil, to add up to 53 proteins. To estimate the actual concentration of each protein, its spectrum is scaled by a concentration factor, and the data with the minimum NRMSD value coupled with visual inspection of predicted spectrum overlay experimental spectrum is to be judged as best fit. SSNN3, a concentration scaling module, was used to automate this process for each protein spectrum with which the effect of temperature or solvent can be identified. SSNN estimated the actual protein concentration and its structure vectors for the following proteins, membrane peptide: α -helix, polylysine: β -sheet to be 0.95 (0.83,0,0.05,0.12), 0.9 (0.07,0.44,0.20,0.29), for insulin at room temperature the estimated secondary structure exhibits about 9% lower helicity relative to the crystal structure. Other examples of proteins this concentration approach was applied to are lipoprotein, ZapA (wild and mutant), and toxins [24].

In 2020, Bansal *et al.* [34] demonstrated the use of SOMSpec in predicting the secondary structure of solid-binding peptide composites. The CD spectra of these special materials were derandomised by removing different fractions of random coil, RC, after which SOMSpec was ran with a 50 x 50 map, 5 BMUs, etc., to enable identification of best fit. It was evident that adding RC caused core protein to lose folded structure, with the best fit obtained when RC accounted for approximately 14% of the PG residues, including the linker. A year later, Adewale *et al* automated the *ad hoc* approach for derandomisation of CD spectra of protein or peptide using MATLAB GUI. This approach was applied to BSA, lysozyme, insulin and uperin U3.x peptides as function of temperature and, or addition of buffer, with best fit determine for each condition. It appeared to be straightforward for BSA and most uperin U3.x peptides, contrary, the best fits were not obvious for lysozyme and insulin [21,35].

SOMSpec was first used to extract secondary structure information from the IR absorbance data by Corujo *et al* in 2018 [36]. They used reference set from BioTools to obtain structure estimates for hemoglobin, lysozyme and concanavalin collected with ATR, and transmission spectroscopy techniques. The results for these techniques were somehow comparable to those of the crystal structures. Subsequently, Corujo *et al*, undertook LOOCV with the aim of evaluating the reliability and accuracy of the secondary structure prediction. It was found that reference proteins with high helix contributed disproportionately to the average prediction errors for the two notable structures in aqueous protein samples, while applying fourier self-deconvolution, FSD to the spectra did not enhance the structure estimates for notable protein secondary structures [22]. Most recently, SOMSpec was applied to predict the secondary structure of Raman protein spectra by LOOCV strategy. The prediction errors for both aqueous and solid-state proteins' spectra from the crystal structures for α -helix and β -sheet were below 10% except for a few solid-state proteins. The obtained estimates from SOMSpec compared well with partial least squares analysis, PLS predictions and performed far better than band fitting for aqueous state, however, reverse is true for solid state proteins [23].

SOMSPEC Future

SSNN has evolved into a more robust structure algorithm named SOMSpec, which accept proteins' spectra from three optical spectroscopies as input. Over the past decade, this application has

provided answers to problems like estimating the true concentration of sample protein in a solution and its corresponding secondary structure, as well as predicting the secondary structure of proteins with significant disordered content. It is noteworthy that, across a range of applications, its performance has been shown to compare well with established statistical methods such as SELCON-3. It is anticipated that extending this structure algorithm to IR and Raman spectra to tease out structural information will enhance its popularity among spectroscopists and within the biopharmaceutical industry. Moreover, SOMSpec's flexibility with respect to wavelength range and secondary structure assignment may further enable its extension to diverse applications, involving reference and test with any form of attributes, thereby opening new avenues to address additional structural problems along with other factors in the years to come.

CURRENT CHALLENGES AND LIMITATIONS OF SSNN AND SOMSPEC

Despite their demonstrated utility, SSNN and SOMSpec face several challenges that may limit broader adoption and application. A major limitation is their dependence on the quality and diversity of the reference datasets used during training. Proteins containing rare folds, extensive intrinsic disorder, membrane-spanning domains may not be adequately represented, potentially reducing prediction accuracy for such systems [19-21,24,25].

The predictive performance of SOM algorithms also depends on optimization of parameters such as map size, neighborhood radius, number of iterations, learning-rate function, and best matching units (BMUs) [20,31,32]. Although validation studies have identified parameter ranges that generally generate reliable predictions, achieving optimal performance may still require considerable user experience and computational experimentation.

Finally, adoption of SOM-based algorithms remains lower than that of more established methods such as CDSSTR, SELCON3, and CONTIN [13,37]. Wider dissemination of user-friendly software, larger benchmarking studies, and standardized validation procedures will be necessary to facilitate broader acceptance within the structural biology and spectroscopy communities.

CONCLUSIONS

Self-organizing map methodologies have contributed significantly to the interpretation of protein spectroscopic data over the past three decades. Beginning with K2D and SOMCD and progressing through SSNN to SOMSpec, these approaches have consistently demonstrated the ability to extract meaningful secondary-structure information from spectroscopic datasets [16,17, 20, 25].

SSNN introduced a flexible SOM-based framework for the analysis of circular dichroism spectra, whereas SOMSpec extended this capability to infrared and Raman spectroscopy, thereby broadening the range of spectroscopic applications that can be addressed using a common machine-learning platform [20,22, 23]. Validation studies have demonstrated that these methods provide reliable structural estimates and can address specialized problems such as concentration uncertainty and characterization of proteins with substantial disordered content [21-23,24, 35, 36].

Although challenges remain regarding reference dataset diversity, parameter optimization, and wider adoption, SOMSpec represents a versatile and promising platform for spectroscopic structure determination. Continued methodological advancement and integration with modern artificial intelligence technologies are expected to further enhance its impact on structural biology, biophysics, and biopharmaceutical research.

ACKNOWLEDGEMENT

We gratefully acknowledge the insightful and thoughtful discussions with Alison Rodger on self-organizing map for determination of secondary structure of proteins/peptides. Adewale Olamoyesan also would like to acknowledge Anothly Nash for his invaluable suggestions on explicating self-organizing map methodology within some chapters in Olamoyesan's doctoral dissertation, as Nash's comments were instrumental in conceiving the title and outline of this minireview

REFERENCES

1. Minor, D. & Kim, P. (1996). Context-dependent secondary structure formation of a designed protein sequence. *Nature*, 380, 730–734.
2. Wright, P. E. & Dyson, H. J. (1999). Intrinsically unstructured proteins: Re-assessing the protein structure–function paradigm. *Journal of Molecular Biology*, 293, 321–331.
3. Callaway, E. (2015). The revolution will not be crystallized: A new method sweeps through structural biology. *Nature*, 525, 172.
4. Cavanagh, J., Fairbrother, W. J., Palmer, A. G., III & Skelton, N. J. (2007). Protein NMR spectroscopy: Principles and practice (2nd ed.). Academic Press.
5. Orlov, I., et al. (2012). 4.5 Å-resolution cryo-EM structure of the large ribosomal subunit of *Tetrahymena thermophila*. *Nature*, 483, 127–132.
6. Johnson, W. C., Jr. (1988). Secondary structure of proteins through circular dichroism spectroscopy. *Annual Review of Biophysics and Biophysical Chemistry*, 17, 145–166.
7. Pelton, J. T. & McLean, L. R. (2000). Spectroscopic methods for analysis of protein secondary structure. *Analytical Biochemistry*, 277, 167–176.
8. Kong, J. & Yu, S. (2007). Fourier transform infrared spectroscopic analysis of protein secondary structures. *Acta Biochimica et Biophysica Sinica*, 39, 549–559.
9. Berjot, M., Marx, J. & Alix, A. J. P. (1987). Determination of the secondary structure of proteins from the Raman amide I band: The reference intensity profiles method. *Journal of Raman Spectroscopy*, 18, 289–300.
10. Provencher, S. W. & Gloeckner, J. (1981). Estimation of globular protein secondary structure from circular dichroism. *Biochemistry*, 20, 33–37.
11. van Stokkum, I. H., Spoelder, H. J., Bloemendal, M., van Grondelle, R. & Groen, F. C. (1990). Estimation of protein secondary structure and error analysis from circular dichroism spectra. *Analytical Biochemistry*, 191, 110–118.
12. Compton, L. A. & Johnson, W. C., Jr. (1986). Analysis of protein circular dichroism spectra for secondary structure using a simple matrix multiplication. *Analytical Biochemistry*, 155, 155–167.
13. Sreerama, N. & Woody, R. W. (2000). Estimation of protein secondary structure from circular dichroism spectra: Comparison of CONTIN, SELCON, and CDSSTR methods with an expanded reference set. *Analytical Biochemistry*, 287, 252–260.

14. Spencer, S. E. & Rodger, A. (2021). Bayesian inference assessment of protein secondary structure analysis using circular dichroism data. *Analytical Methods*, 13, 359–368.
15. Kohonen, T. (1982). Self-organized formation of topologically correct feature maps. *Biological Cybernetics*, 43, 59–69.
16. Andrade, M. A., Chacón, P., Merelo, J. J. & Morán, F. (1993). Evaluation of secondary structure of proteins from UV circular dichroism spectra using an unsupervised learning neural network. *Protein Engineering*, 6, 383–390.
17. Louis-Jeune, C., Andrade-Navarro, M. A. & Perez-Iratxeta, C. (2012). Prediction of protein secondary structure from circular dichroism using theoretically derived spectra. *Proteins*, 80, 374–381.
18. Unneberg, P., Merelo, J. J., Chacón, P. & Morán, F. (2001). SOMCD: Method for evaluating protein secondary structure from UV circular dichroism spectra. *Proteins*, 42, 460–470.
19. Hall, V., Nash, A. & Rodger, A. (2014). SSNN: A neural network method for protein secondary structure fitting using circular dichroism data. *Analytical Methods*, 6, 6721–6726.
20. Ang, D. L. (2019). Biophysical and computational studies of biomacromolecular systems [Doctoral dissertation, Western Sydney University].
21. Olamoyesan, A., Ang, D. & Rodger, A. (2021). Circular dichroism for secondary structure determination of proteins with unfolded domains using a self-organising map algorithm SOMSpec. *RSC Advances*, 11, 23985–23991.
22. Corujo, P. M., Olamoyesan, A., Tukova, A., Ang, D., Goormaghtigh, E., Peterson, J., Sharov, V., Chmel, N. & Rodger, A. (2022). SOMSpec as a general purpose validated self-organising map tool for rapid protein secondary structure prediction from infrared absorbance data. *Frontiers in Chemistry*, 9, 784625
23. Corujo, P.M., Michal, P., Ang, D., Vivian, L., Chmel, N. & Rodger, A. (2025). Prediction of secondary structure content of proteins using Raman spectroscopy and self-organizing maps. *Applied Spectroscopy*, 79(10), 1497–1507.
24. Hall, V., Sklepari, M. & Rodger, A. (2014). Protein secondary structure prediction from circular dichroism spectra using a self-organizing map with concentration correction. *Chirality*, 26, 471–482.

25. Hall, V., Nash, A., Hines, E. & Rodger, A. (2013). Elucidating protein secondary structure with circular dichroism and a neural network. *Journal of Computational Chemistry*, 34, 2774–2786.
26. Aupetit, M., Massotte, P. & Couturier, P. (1999). C-SOM: A continuous self-organizing map for function approximation. In *Proceedings of Intelligent Systems Control (ISC)*.
27. Takatsuka, M. (2001). An application of the self-organizing map and interactive 3-D visualization to geospatial data. In *Proceedings of the 6th International Conference on GeoComputation* (pp. 24–26).
28. Ito, M., Miyoshi, T. & Masuyama, H. (2000). The characteristics of the torus self-organizing map. In *Proceedings of the 6th International Conference on Soft Computing (IIZUKA)* (pp. 239–244).
29. Zhang, J. (2007). *Visualization for information retrieval*. Springer.
30. Grajiarová, L., Mareš, J., Dvořák, P. & Procházka, A. (2012). Biomedical image analysis using self-organizing maps. In *Proceedings of the Annual Conference of Technical Computing*.
31. Natita, W., Wiboonsak, W. & Dusadee, S. (2016). Appropriate learning rate and neighborhood function of self-organizing map for specific humidity pattern classification. *International Journal of Modeling, Optimization and Engineering*, 6, 61–66.
32. Hall, V. A. (2014). Self-organising map machine learning approach to pattern recognition [Doctoral dissertation, University of Warwick].
33. Rahmatbakhsh, M., Gagarinova, A. & Babu, M. (2021). Bioinformatic analysis of temporal and spatial proteome alterations during infections. *Frontiers in Genetics*, 12, 1155.
34. Bansal R., Elgundi Z., Goodchild S.C., Care A, Lord M.S., Rodger A. & Sunna A. (2020). The effect of oligomerization on a solid-binding peptide binding to silica-based materials. *Nanomaterials*, 10, 1070.
35. Olamoyesan, A. & Rodger, A. (2024). Application of derandomisation to Uperin 3.x peptides circular dichroism spectra. *South African Journal of Chemistry*, 78, 67–75.
36. Corujo, M. P., Sklepari, M., Ang, D. L., Millichip, M., Reason, A., Goodchild, S. C., Wormell, P., Amarasinghe, D. P., Lindo, V., Chmel, N. P. & Rodger, A. (2018). Infrared absorbance spectroscopy of aqueous proteins: Comparison of transmission and ATR data collection and analysis for secondary structure fitting. *Chirality*, 30(8), 957–965.
37. Whitmore, L. & Wallace, B. A. (2004). DICHROWEB, an online server for protein secondary structure analyses from circular dichroism spectroscopic data. *Nucleic Acids Research*, 32, W668–W673.