# Neural Network Pairwise Interaction Fields for protein model quality assessment

Alberto J. M. Martin, Alessandro Vullo, Gianluca Pollastri
School of Computer Science and Informatics and
Complex and Adaptive Systems Laboratory
University College Dublin, Belfield, Dublin 4, Ireland

**Abstract**

We present a new knowledge-based Model Quality Assessment Program (MQAP) at the residue level which evaluates single protein structure models. We use a tree representation of the $C_\alpha$ trace to train a novel Neural Network Pairwise Interaction Field (NN-PIF) to predict the global quality of a model. We also attempt to extract local quality from global quality. The model allows fast evaluation of multiple different structure models for a single sequence. In our tests on a large set of structures, our model outperforms most other methods based on different and more complex protein structure representations in both local and global quality prediction. The method is available upon request from the authors. Method-specific rankers may also built by the authors upon request.

## 1 Introduction

In order to use a 3D model of a protein structure we need to know how good it is, as its quality is proportional to its utility [8]. Moreover most protein structure prediction methods produce many reconstructions for any one protein and being able to sift out good ones from bad ones is often the key to the success of a method.

Several different potential or (pseudo-)energy function have been developed with the aim of mapping a three-dimensional structure into its "goodness" or "native-likeness". These can be roughly divided into physics-based and knowledge-based, with some amount of overlap. The former are based on physico-chemical calculations [7, 23, 24, 29, 38], while the latter are based on statistics obtained from a training set of known structures (also termed "decoys") typically generated by computational methods. These structures or decoys can be represented in different ways, embedded or not into a 3D lattice, in order to reduce the conformational space, and potentials can rely on more or less detailed representations of a structure. As simplicity in the representation increases (e.g. if each residue is represented as a single point or sphere), evaluation speed increases since fewer interactions have to be taken into account. By contrast the reliability of the evaluation tends to decrease when details are stripped from a structure [6, 13, 28, 52]. However, there may be other advantages

in the use of simplified representations, such as reduced sensitivity to small perturbations in conformations and the ability to take into account complex effects that cannot be described separately [26], plus a reduced sensitivity to inaccuracies and uncertainties in the data (crystal contacts, high R-factors, etc.).

Knowledge-based potentials are not only used to rank models, but for other aims such as to drive 3D predictions and for model refinement [3,23,24,41,51]; fold recognition [35]; to place side chain and backbone atoms [1,5]; as mentioned, to select native structures from sets of decoys [6, 13]; to predict stability of proteins [20, 21, 57]; or even to predict residue residue contact maps [31, 39, 46].

Several different representations of protein structures have been used in knowledge-based potentials. They can be classified in: 1) single point/sphere representations of each residue; 2) two or more points for each residue; 3) full atom models. In the first group each residue is usually represented by its $C_\alpha$ atom [12, 19, 22, 27, 34, 42] or by their $C_\beta$ atom [4,6,10,13,21,35,42], pseudo-$C_\beta$, side chain (SD) centre of mass or SD centroids [54]. In group 2) each residue is represented by two or more points, but fewer than its number of atoms; these points range from only two (e.g. $C_\alpha$+$C_\beta$, $C_\alpha$+pseudo-$C_\beta$ [14]), backbone atoms and some sort of representation of side chains like the $C_\beta$ [57], different rigid-body blocks/fragments [28]. In group 3) are many different potentials [4, 6, 11, 20, 37, 47, 48, 57]. These depend on different reference states and on different physico-chemical assumptions, whereas others are learnt from examples, typically via Machine Learning algorithms [19, 33, 48, 50].

There is a further group of potential functions - those based on clustering of many different structure predictions, and consensus methods. These usually outperform single model evaluation methods [8, 17, 33, 49]. Consensus methods make use of several different potentials to know the quality of the prediction [33, 50, 51]. The potentials are joined in a (possibly weighted) average. Clustering methods use similarities among high numbers of models obtained either from different methods (in this case they are hard to apply outside the CASP experiment environment($http://predictioncenter.org/$), as many of the methods and their predictions are not easily available outside CASP), or to rank high numbers of reconstructions made by a single method [36, 48, 56].

The method presented here only uses information obtained from the $C_\alpha$ trace and the sequence of residues associated to it. The main advantage is that the quality of a model is assessed based on its overall topology, rather than based on local details. Moreover, there is no need to model backbone and side chain atoms before evaluating a structure, which allows many more $C_\alpha$ traces with different conformations to be produced. From $C_\alpha$ traces it is possible to model backbone and side chain atoms fairly accurately [1, 5], but this may be more computationally expensive than predicting several conformations of the protein structure as simple $C_\alpha$. If $C_\alpha$ traces can be evaluated effectively, backbone and side chains may be modelled only for those that are deemed to be accurate.

## 2 Methods

Protein model quality is often measured as the scaled distance between $C_\alpha$s of models to their postions in the native structure after optimal superimposition of the structures. Here we encode information obtained solely from the $C_\alpha$ trace. First we represent the $C_\alpha$ trace of each structure model as a directed acyclic graph (rooted tree), in which the outer nodes are pairwise interactions. Each residue in the $C_\alpha$ is encoded into a vector describing its environment. Interactions among $C_\alpha$s are simply characterised by their distance and angles between pseudo-$C_\beta$ , alongside the two vectors encoding the residues involved. Environments are described by several angles, distances among neighbours, pseudo-Solvent Accessibility (SA), and coarse packing information. Both interactions and environment descriptors are described in section 2. All these numerical descriptors are computed from the $C_\alpha$ trace and are fed into a model (Neural Network Pairwise Interaction Field, NN-PIF) trained to predict global quality. In the NN-PIF each $C_\alpha$ (i.e. its interactions with all the other $C_\alpha$s) is mapped into a hidden state, which contains the contribution of that residue to the global quality of the structure. Two $C_\alpha$s are considered as interacting if they are closer than a fixed distance threshold (we use 20Å, see 2.3). The hidden vectors for all $C_\alpha$ are then combined and mapped to a global quality measure. The NN-PIF allows us to evaluate all the interactions at the same time, whereas other knowledge based potentials generally evaluate interactions separately. To train the NN-PIF we use models submitted to previous CASP editions [2], as the main purpose of this MQAP is to rank models from different prediction systems. No native structures are included in the training set.

Incidentally, the NN-PIF is able to evaluate all the 3D server models submitted to CASP7 [2], as it only depends on the $C_\alpha$ trace and a number of predictors only submitted this. Our system runs on all the CASP7 server models (about 24000, including AL models - see section 2.6) in approximately 1 hour on a PentiumIV 3 GHz processor. However, given the very large amount of weight sharing, we found that training has to proceed very slowly, and training times can be dire (in the region of months on a single CPU).

### 2.1 Graph representation of a protein structure and NN-PIF

Ways to represent structured information (in the form of a DAG) by recursive neural networks have been described in the past (e.g. [15, 16, 43]), and training can proceed by extensions of the backpropagation algorithm. In our case the complex of interactions among all $C_\alpha$ atoms may be naturally represented as an undirected graph in which nodes are atoms and labelled edges describe the nature of the interactions. However, in order to represent a structure as a DAG, we consider interactions themselves as nodes (see figure 1). If the identity and environment of the $i^{th}$ atom is encoded by a vector $a_i$, and the interaction between the $i^{th}$ and $j^{th}$ atom is described by vector $d_{ij}$, then each pair of atoms is mapped into a hidden state $X_{ij}$ as:

$$X_{ij} = F(a_i, a_j, d_{ij}) \tag{1}$$

The function $F()$ is implemented by a feed-forward neural network with a single hidden layer and a linear output. The hidden states are then combined together for each residue $i$, yielding a hidden state for the whole residue:

$$Y_i = \sigma(K \sum_{j \in Ci} X_{ij}) \tag{2}$$

where $Ci = \{\forall j | d(a_i, a_j) < 20\text{Å}\}$, $d()$ is the Euclidian distance, and $K$ is a normalisation constant. Finally, the hidden states for all residues are averaged into a single output, which represents a single property for the whole structure:

$$O = \frac{1}{L} \sum_{i=1}^{L} Y_j \tag{3}$$

The function $F()$ is assumed to be stationary, hence the same network is replicated for all the interactions. The overall NN-PIF architecture is trained by gradient descent. We assume the error to be the squared difference between the network output and the desired property (in our case the "goodness" of the structure). The gradient can be easily computed in closed form, via a version the backpropagation algorithm. It should be noted that during training the gradient is computed for each replica of the network $F()$, hence there will be as many partial derivatives of the global error with respect to each free parameter in $F()$ as there are interacting pairs of atoms in a model. The contributions to the gradient from each replica of $F()$ are added up component by component to yield the final gradient.

Also notice how here the states describing a pairwise interaction and all the interactions of a residue ($X_{ij}$ and $Y_i$, respectively) are mapped into the output $O$ through a fixed function with no free parameters. It is also possible to devise a model in which $X_{ij}$ and $Y_i$ are vectors and the average, or sum, of all $Y_i$ is mapped into the desired output through a further feed-forward network. We are currently implementing such model.

## 2.2 NN-PIF configuration and training

The NN-PIF we train here has 10 hidden neurons in the hidden layer of the network implementing $F()$. The learning rate is set to 3.0. During training, the weights are updated after the gradient for all pairs of residues of a single complete structure model has been computed. As we were participating to the $7^{th}$ edition of the CASP experiment during the preparation of this work, it was possible to run comparisons against other methods in real time. Our performances were evaluated on the models submitted by servers at CASP (see below for details).

## 2.3 NN-PIF inputs

To describe two residues in contact ($i$ and $j$), their environments and the interaction between them, the following descriptors are used:

- Spatial neighbours: All those residues whose $C_\alpha$ is closer than $20$Å to the $C_\alpha$ of residue $i$ are considered neighbours of $i$. Although beyond $15$Å the effect of an atom on another is generally considered negligible, it should be noted that we do not aim at obtaining a physical model.

- Local backbone conformation. We do not explicitly use secondary structure, unlike other MQAP methods [4, 48]. Instead we use several structure descriptors computed from the $C_\alpha$ trace:

  - Distances between all $C_\alpha$ in $[i-2, i+2]$ to each other. A smaller set of distances was used in [32] and to validate protein models in [22].

  - $C\alpha_{i-1} - C\alpha_i - C\alpha_{i+1}$ angles and dihedral angles (angles between vectors formed by $\overrightarrow{C\alpha_{i-1} - C\alpha_i}$ and $\overrightarrow{C\alpha_i - C\alpha_{i+2}}$, and between vectors formed by $\overrightarrow{C\alpha_i - C\alpha_{i+1}}$ and $\overrightarrow{C\alpha_i - C\alpha_{i+2}}$ for both $i$ and $j$, as in [22, 25, 30].

  - Distance to sequence neighbours: distances between each $C_\alpha$ in $[i-2, i+2]$ and each $C_\alpha$ in $[j-2, j+2]$. A smaller set of distances was used in [32] to assign $\beta$ Sheets.

  - Relative spatial orientation with respect to sequence neighbours: angles between pseudo-$C_\beta$ vectors (placed in the direction of the vector formed by the sum of $\overrightarrow{(i-1, i)}$ and $\overrightarrow{(i, i+1)}$ - this vector is also used to compute pseudo-solvent accessibility, see below). Angles between pseudo-$C_\beta$ of each residue in $[i-1, i+1]$ and in $[j-1, j+1]$ against all the other pseudo-$C_\beta$ vectors of the residues in the same ranges.

- Residue identities: both residues in contact ($i$ and $j$) are one-hot encoded (20 inputs/residue).

- Pseudo-Solvent Accessibility (SA) as HSE measure [18]. Briefly, a sphere of radius $6.5$Å centred on a $C_\alpha$ is divided into two hemispheres by the plane whose normal vector is the sum of vectors $\overrightarrow{C\alpha_{i-1} - C\alpha_i}$ and $\overrightarrow{C\alpha_i - C\alpha_{i+1}}$, then the number of other atoms falling in either hemisphere is counted.

- Coarse Packing (HSE8): we further split the a sphere, this time into $8$ slices induced by three perpendicular planes and count the number of $C_\alpha$s within each slice. The normal vectors of these three planes are: the pseudo-$\beta$ vector; the vector obtained from the cross-product of $\overrightarrow{C\alpha_{i-1} - C\alpha_i}$ and $\overrightarrow{C\alpha_i - C\alpha_{i+1}}$; the cross-product of the first two. This time we use a sphere of $13$Å .

Some of these inputs can not be computed for residues at the protein termini, in which case they are set to 0. If chain breaks occur ($C_\alpha$s separated by more than $4.7$Å), residues at the edges of the break are treated as termini.

## 2.4 Model quality measurement: structural distance

All the methods we consider are based on the Euclidean distance between atoms in the model and in the native protein structure after structural superimposition. Those described below are or may be computed using $C_\alpha$ traces alone. We choose the last in the list (TM score) as our target function.

- RMSD: Root Mean Square Distance. Widely used in the field. Not very suitable to be used as desired output because above a certain threshold it becomes insensitive to highly similar substructures when many atoms of the model are wrongly modelled; also, it is not constrained to a fixed range.
- GDT_TS [53]: it identifies maximum common substructures based on several distance thresholds (e.g. $1$, $2$, $4$ and $8$ Å as used in CASP). It may miss fine details because all the atoms within a range (e.g. $(4, 8]$) contribute the same to the scoring function and all the residues without coordinates (atoms not present in the model), contribute the same as those residues further than the highest distance threshold.
- MaxSub score: the average scaled distance using a maximum distance threshold of 3.5 Å (residues further than 3.5 Å do not contribute to the score [40]).
- Average S score: the average scaled distance using a distance threshold of 5Å [9].
- TM-Score [55]: based on scaled RMSD but TM-Score is scaled by the protein length - each residue in the model and native structure contributes to it. It is a number in the $[0, 1]$ range allowing its direct use as desired output.

## 2.5 Performance Test and measures

Several different potential quality measures can been used, and here we will use the following ones:

- Enrichment ($E15\%$): the number of top $15\%$ models found among the top $15\%$ top ranked models, divided by the number obtained in a random selection ($15\%$ x $15\%$ x number of structures in the model set [45].
- Pearson correlation coefficient ($r$) for both global quality and local quality.
- Spearman's rank correlation coefficient ($\rho$).
- Recall ($R = \frac{TP}{TP+FN}$, where $TP$ are true positives, $FN$ false negatives). We use Recall to measure the quality of the output after quantisation into categories (e.g. good vs. bad, defined as greater or smaller than a given threshold).
- Precision ($P = \frac{TP}{TP+FP}$, where $FP$ are false positives). We use Precision on quantised output/target as for Recall.

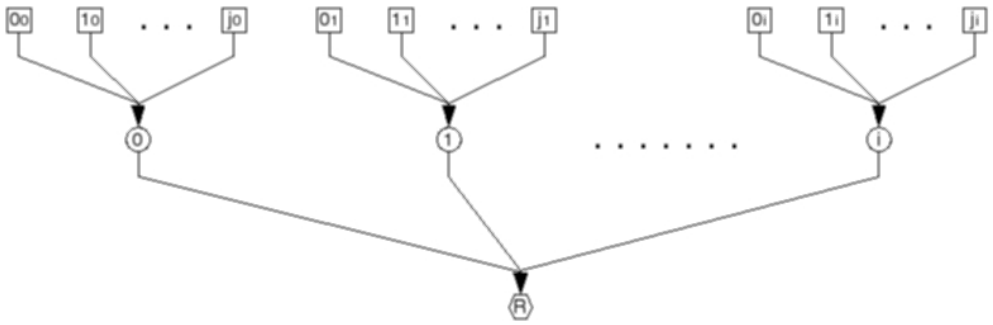


Figure 1: Tree representation of a $C_\alpha$ trace. Squares, neighbour nodes, are connected to the inner nodes, circles. All the inner nodes, one for each residue in the $C_\alpha$ trace, are connected to the root node represented as an hexagon.

### 2.6 Dataset

The main purpose of this MQAP is to know how good protein structure models are (how close are they to the native structure), sorting them by their quality. Because of this we train the NN-PIF only on models submitted to CASP5, 6 and 7, while no native structure is used for training.

To train the NN-PIF we use all human and server predictions of targets for which all 3D predictions were assessed. We randomly divide these into 5 folds, each of which is used to train a different model. To avoid redundancies in the training data (i.e. many models with almost exactly the same $C_\alpha$ trace), TM-Score is divided into $0.001$ size bins and only one model per bin for each target is included into each training fold.

All the methods are evaluated on the CASP7 [2] server predictions. We present results on all targets, and on two slightly smaller sets of targets, in order to allow direct comparisons with other methods on the same sets. When testing on one protein, we ensemble those networks that were trained on folds not containing it. This is similar to an n-fold cross validation, except that CASP5 and CASP6 proteins are only included in training sets.

## 3 Discussion

We compare our MQAP with other methods for global model quality assessment, and also show results for single model quality. We refer to our method as DISTILLF (its identifier at CASP8). To gauge other methods' performances we look at published results. A number of methods were tested on CASP7 targets, and for these the results are extracted from the CASP web site. The results for other methods are quoted from their respective publications. Table 1 is extracted from [33], and shows average per target Spearman's rank correlations for a number of methods. The table is computed on 87 CASP7 targets. ModFOLD is a

consensus method and 3D-Jury is a clustering method, i.e. they are not primary methods, but rather rely on multiple other methods for their predictions. As such, they are in a category of their own and it is not fair to compare them with single model methods. All the other methods evaluate single models, and DISTILLF clearly outperforms all of them, has a rank correlation 7.4% higher than the next method based on TM-Score, and 2.2% based on GDT_TS. The smaller gain is expected as DISTILLF is trained to predict TM-Score.

Table 2 is extracted from [4], and compares QMEAN with other single model evaluation methods available. QMEAN, as DISTILLF, was developed after CASP7. This table is computed on the 95 targets evaluated at CASP7 and on all those server predictions for which all the programs compared were able to make a prediction (22427 models in total). It should be noted that the table is based on GDT_TS, while DISTILLF is trained to predict TM-score, and as such it is at an obvious disadvantage. In the table we also report DISTILLF results based on TM-Score. Not unexpectedly, its performance increases in this case. DISTILLF performs well on correlation measures, and is only outperformed by the most complete QMEAN potentials, which are based on full-atom models, while DISTILLF only relies on $C_\alpha$ traces, or on a number of interactions smaller by two orders of magnitude. However DISTILLF is less than perfect at selecting the best models ($E_{15\%}$ measure), indicating that it is better at estimating the absolute quality of a model, than at ranking models that are very similar. This is not surprising, as very similar models (e.g. good predictions based on homology to known structures) often have to be distinguished based on local atomic details, which DISTILLF does not rely on.

To measure the performance of DISTILLF on single residue quality, we use all the TS and AL models submitted by automatic servers for all the 98 CASP7 targets. All measures are computed on 4.55 million residues. S and MaxSub scores are generated fixing $d_0$ in the TM-Score package to 5 and 3.5Å respectively. We take the $Y_i$ hidden value from the NN-PIF (see eq.2) as the local estimate of quality for residue $i$ by DISTILLF. Table 3 shows the Pearson's correlation against TM, S and MaxSub scores. Table 4 reports results on identifying well modelled residues (those with a scaled distance $\geq 0.7$). Table 4 reports results on identifying wrongly modelled residues (those with an actual scaled distance $\leq 0.3$).

# 4 Conclusion

In this manuscript we have described a novel predictor of model quality for protein structure prediction. The main novelty of the predictor is the model it is based on, a neural network designed to estimate properties of sets of pairwise interactions, which we have provisionally termed NN-PIF. A single feed-forward network is used to estimate each of the interactions, and the outputs of all replicas of the network are combined without resorting to any free parameter, to yield a single property. In this manuscript we predict the "goodness" or "native-likeness" of a protein model. However, the model can be used more in general to learn about data that can be represented as undirected graphs. While the model we used in this manuscript is fairly simple, it can be easily extended to more expressive versions,

| $Method$ | $TM-Score$ | $GDT\_TS$ |
|---|---|---|
| 3D-Jury [33] | 0.87 | 0.857 |
| ModFOLD [33] | 0.732 | 0.754 |
| PROQ [48] | 0.574 | 0.587 |
| Pcons [50] | 0.557 | 0.58 |
| ProQ-MX [47] | 0.55 | 0.556 |
| ModSSEA [33] | 0.506 | 0.52 |
| MODCHECK [35] | 0.412 | 0.444 |
| ProQ-LG [47] | 0.289 | 0.326 |
| DISTILLF | 0.647 | 0.609 |

Table 1: Spearman's rank correlation on CASP7 targets server submitted models for the 87 targets evaluated in [33], both TS and AL models used.

| $Method$ | $r^2$ | $\rho$ | $E_{15\%}$ |
|---|---|---|---|
| Modcheck [35] | 0.64 | 0.59 | 2.7 |
| RAPDF [37] | -0.5 | 0.5 | 2.44 |
| DFIRE [57] | -0.39 | 0.53 | 2.59 |
| ProQ [47] | 0.36 | 0.26 | 1.22 |
| ProQ_SSE [47] | 0.54 | 0.43 | 1.71 |
| FRST [44] | -0.57 | 0.53 | 2.36 |
| QMEAN3 [4] | -0.65 | 0.58 | 2.57 |
| QMEAN4 [4] | -0.71 | 0.63 | 2.76 |
| QMEAN5 [4] | -0.72 | 0.65 | 2.9 |
| DISTILLF_GDT | 0.65 | 0.59 | 2.34 |
| DISTILLF_TMS | 0.68 | 0.64 | 2.53 |

Table 2: All CASP7 server models for the 95 targets evaluated. Average per target. All the methods were evaluated on 22427 models of the 95 targets used in [4]. Other methods extracted from [4].

| $Method$ | $r$ |
|---|---|
| DISTILLF_TMS | 0.71 |
| DISTILLF_MaxSub | 0.70 |
| DISTILLF_S | 0.69 |

Table 3: Single residue correlation coefficient on all the CASP7 targets server models (4.55 million residues, both AL and TS models).

| $ScaledDistance$ | $TP$ | $FP$ | $TN$ | $FN$ | $P$ | $R$ |
|---|---|---|---|---|---|---|
| $TM-Score$ | 0.963 | 0.396 | 1.777 | 1.417 | 0.708 | 0.405 |
| $MaxSub$ | 0.926 | 0.433 | 1.846 | 1.348 | 0.681 | 0.407 |
| $Sscore$ | 0.815 | 0.544 | 2.073 | 1.121 | 0.600 | 0.421 |

Table 4: Ability to identify correctly modelled residues. Correct residues are those with scaled distance $\geq 0.7$. TP, TN, FP and FN are, respectively, true positives, true negatives, false positives and false negatives, in millions of residues. True values obtained with TM-Score package with default options, and setting $D_0$ to 3.5Å and 5Å for the MaxSub and S scores respectively. $4.55 \cdot 10^6$ residues in total.

| $ScaledDistance$ | $TP$ | $FP$ | $TN$ | $FN$ | $P$ | $R$ |
|---|---|---|---|---|---|---|
| $TM-Score$ | 1.107 | 1.543 | 1.522 | 0.381 | 0.418 | 0.744 |
| $MaxSub$ | 1.144 | 1.506 | 1.492 | 0.411 | 0.432 | 0.736 |
| $Sscore$ | 1.361 | 1.29 | 1.376 | 0.526 | 0.513 | 0.721 |

Table 5: Ability to identify badly modelled residues. Badly modeled residues are those with scaled distance $\leq 0.3$. TP, TN, FP and FN are, respectively, true positives, true negatives, false positives and false negatives, in millions of residues. True values obtained with TM-Score package with default options, and setting $D_0$ to 3.5Å and 5Å for the MaxSub and S scores respectively. $4.55 \cdot 10^6$ residues in total.

for instance one in which each pairwise interaction is mapped into a hidden vector (rather than a single hidden state, as it is in this work), and a combination of all hidden vectors is then mapped to a property of interest via a further network. We are currently working on such model. Another simple extension is one in which properties of single nodes of the undirected graph are predicted. This can be achieved by mapping hidden states describing each single node ($Y_i$ in eqn.2, or a multi-dimensional extension thereof) to the property of the node via a second network.

The DISTILLF predictor which we have described in this manuscript, relies on simple $C_\alpha$ traces as inputs. The fact that we can use such simple representation induces a set of interactions that is two orders of magnitude smaller than that of a full-atom model, allows very large scale processing of protein models, and is a direct consequence of using a model that does not rely on physico-chemical laws, but only on geometrical information and machine learning. In spite of its simplicity, we have shown that DISTILLF is accurate, more so than any of the CASP7 primary algorithms for model quality assessment, and only slightly less accurate than a newer, far more computationally complex system based on full-atom models. Although DISTILLF is meant to predict global model quality, we have also shown that an accurate estimate of local quality can be extracted very simply from it.

It is also important to note that, although much of the appeal of DISTILLF is that it relies on $C_\alpha$ traces, the NN-PIF model is equally suited to deal with full-atom representations of

molecules. We are currently testing NN-PIF on the prediction of protein-ligand binding energies based on full-atom models, with encouraging preliminary results.

A further future/current direction of research is whether NN-PIF may be applied directly as potentials for the *ab initio* prediction of protein structures. In this case, rather than on endpoints of structure prediction searches, decoys representing intermediate stages of the search need to be used for training. Although building sets of examples with the correct distribution may be a hard task, and so is training a network on a potentially enormous set, even limited success at this task may yield a fast, flexible predictor which could be input a vast range of non-homogeneous information.

## Funding

This work is supported by Science Foundation Ireland grant 05/RFP/CMS0029, grant RP/2005/219 from the Health Research Board of Ireland and a UCD President's Award 2004.

## Acknowledgement

We would like to thank Liam McGuffin for his benchmarks [33], and Pascal Benkert for kindly providing us lists of CASP targets.

## References

[1] S.A. Adcock. Peptide backbone reconstruction using dead-end elimination and a knowledge-based forcefield. *J. Comput. Chem.*, 25:16–27, 2004.

[2] J.N. Battey, J. Kopp, L. Bordoli, R.J. Read, N.D. Clarke, and T. Schwede. Automated server predictions in casp7. *Proteins*, 69(Suppl 8):68–82, 2007.

[3] D. Baú, G. Pollastri, and A. Vullo. Distill: a machine learning approach to ab initio protein structure prediction. *Analysis of Biological Data: A Soft Computing Approach, S. Bandyopadhyay, U. Maulik and J. T. L. Wang eds., World Scientific*, 2006.

[4] P. Benkert, S.C.E. Tosatto, and D. Schomburg. Qmean: A comprehensive scoring function for model quality assessment. *PROTEINS: Structure, Function, and Bioinformatics*, 71(1):261–277, 2008.

[5] M.J. Bower, F.E. Cohen, and R.L. Dunbrack. Prediction of protein side-chain rotamers from a backbone-dependent rotamer library: A new homology modelling tool. *J. Mol. Biol.*, 267:1268–1282, 1997.

[6] A. Colubri, A.K. Jha, M. Shen, A. Sali, R.S. Berry, T.R. Sosnick, and K.F. Freed. Minimalist representations and the importance of nearest neighbour effects in protein folding simulations. *J. Mol. Biol.*, 363:835–857, 2006.

[7] W.D. Cornell, P. Cieplak, C.I. Bayly, I.R. Gould, K.M. Merz, D.M. Ferguson, D.C. Spellmeyer, T. Fox, J.W. Caldwell, and P.A. Kollman. A second generation force field for the simulation of proteins, nucleic acids, and organic molecules. *J. Am. Chem. Soc.*, 117:5179–5197, 1995.

[8] D. Cozzetto, A. Kryshtafovych, M. Ceriani, and A. Tramontano. Assessment of predictions in the model quality assessment category. *Proteins*, 69(Suppl 8):175–183, 2007.

[9] S. Cristobal, A. Zemla, D. Fischer, L. Rychlewski, and A. Elofsson. A study of quality measures for protein threading models. *BMC Bioinformatics*, 2(5):15, 2001.

[10] Q. Dong, X. Wang, and L. Lin. Novel knowledge-based mean force potential at the profile level. *BMC Bioinformatics*, 7:324, 2006.

[11] D. Eisenberg, R. Lthy, and J.U. Bowie. Verify3d: assessment of protein models with three-dimensional profiles. *Methods Enzymol.*, 277:396–404, 1997.

[12] Y. Feng, A. Kloczkowski, and R.L. Jernigan. Four-body contact potentials derived from two protein datasets to discriminate native structures from decoys. *PROTEINS: Structure, Function, and Bioinformatics*, 68:57–66, 2007.

[13] J.E. Fitzgerald, A.K. Jha, A. Colubri, T.R. Sosnick, and K.F. Freed. Reduced $c_\beta$ statistical potentials can outperform all-atom potentials in decoy identification. *Protein Science*, 16:2123–2139, 2001.

[14] F. Fogolari, L. Pieri, A. Dovier, L. Bortolussi, G. Giugliarelli, A. Corazza, G. Esposito, and P. Viglino. Scoring predictive models using a reduced representation of proteins: model and energy definition. *BMC Structural Biology*, 7(15):17, 2007.

[15] P. Frasconi. An introduction to learning structured information. *Adaptative Processing LNAI 1387(Springer-Verlag)*, 1998.

[16] P. Frasconi, M. Gori, and A. Sperduti. A general framework for adaptive processing of data structures. *IEEETNN*, 9(5):768–786, 1998.

[17] K. Ginalski, A. Elofsson, D. Fischer, and L. Rychlewski. 3d-jury: a simple approach to improve protein structure predictions. *Bioinformatics*, 19(8):1015–1018, 2003.

[18] T. Hamelryck. An amino acid has two sides: A new 2d measure provides a different view of solvent exposure. *PROTEINS: Structure, Function, and Bioinformatics*, 59:38–48, 2005.

[19] M. Heo, S. Kim, E.J. Moon, M. Cheon, K. Chung, and I. Chang. Perceptron learning of pairwise contact energies for proteins incorporating the amino acid environment. *Phys. Rev. E Stat. Nonlin. Soft Matter Phys.*, 72:011906, 2005.

[20] C. Hoppe and D. Schomburg. Prediction of protein thermostability with a direction- and distance-dependent knowledge-based potential. *Protein Science*, 14:2682–2692, 2005.

[21] J. Khatun, S.D. Khare, and N.V. Dokhlyan. Can contact potentials reliably predict stability of proteins? *J. Mol. Biol.*, 336:1223–1238, 2004.

[22] G.J. Kleywegt. Validation of protein models from c-alpha coordinates alone. *J. Mol. Biol.*, 273:371–376, 1997.

[23] E. Krieger, T. Darden, S.B. Nabuurs, A. Finkelstein, and G. Vriend. Making optimal use of empirical energy functions: Force-field parameterisation in crystal space. *PROTEINS: Structure, Function, and Bioinformatics*, 57:678–683, 2004.

[24] E. Krieger, G. Koraimann, and G. Vriend. Increasing the precision of comparative models with yasara nova a self-parameterising force field. *PROTEINS: Structure, Function, and Bioinformatics*, 47:393–402, 2002.

[25] G. Labesse, N. Colloc'h, J. Pothier, and J.P. Mornon. P-sea: a new efficient assignment of secondary structure from c alpha trace of proteins. *CABIOS*, 13(3):291–295, 1997.

[26] L. Leherte. Application of multiresolution analyses to electron density maps of small molecules: Critical point representations for molecular superposition. *J. of Math. Chem.*, 29(1):47–83, 2001.

[27] C. Loose, J.L. Klepeis, and C.A. Floudas. A new pairwise folding potential based on improved decoy generation and side-chain packing. *PROTEINS: Structure, Function, and Bioinformatics*, 54:303–314, 2004.

[28] M. Lu, A.D. Dousis, and J. Ma. Opuspsp: An orientation-dependent statistical all-atom potential derived from side-chain packing. *J. Mol. Biol.*, 376:288–301, 2008.

[29] A.D. MacKerell, D. Bashford, M. Bellott, R.L. Dunbrack, J.D. Evanseck, M.J. Field, S. Fischer, J. Gao, H. Guo, S. Ha, D. Joseph-McCarthy, L. Kuchnir, K. Kuczera, F.T.K. Lau, C. Mattos, S. Michnick, T. Ngo, D.T. Nguyen, B. Prodhom, W.E. Reiher, B. Roux, M. Schlenkrich, J.C. Smith, R. Stote, J. Straub, M. Watanabe, J. Wiorkiewicz-Kuczera, D. Yin, and M. Karplus. All-atom empirical potential for molecular modelling and dynamics studies of proteins. *J. Phys. Chem.*, 102:3586–3616, 1998.

[30] I. Majumdar, S.S. Krishna, and N.V. Grishin. Palsse: A program to delineate linear secondary structural elements from protein structures. *BMC Bioinformatics*, 6(202):24, 2005.

[31] A.J.M. Martin, D. Baú, I. Walsh, A. Vullo, and G. Pollastri. Long-range information and physicality constraints improve predicted protein contact maps. *Journal of Bioinformatics and Computational Biology*, 6(5), 2008.

[32] J. Martin, G. Letellier, A. Marin, J.F. Taly, Brevern de, Gibrat A.G., and J.F. Protein secondary structure assignment revisited: a detailed analysis of different assignment methods. *BMC Struct. Biol.*, 5:17, 2005.

[33] L.J. McGuffin. Benchmarking consensus model quality assessment for protein fold recognition. *BMC Bioinformatics*, 8:15, 2007.

[34] S. Ngan, M.T. Inouye, and R. Samudrala. A knowledge-based scoring function based on residue triplets for protein structure prediction. *Protein Engineering, Desing & Selection*, 19(5):187–193, 2006.

[35] C.S. Pettitt, L.J. McGuffin, and D.T. Jones. Improving sequence-based fold recognition by using 3d model quality assessment. *Bioinformatics*, 21(17):3509–3515, 2005.

[36] J. Qiu, W. Sheffler, D. Baker, and W.S. Noble. Ranking predicted protein structures with support vector regression. *PROTEINS: Structure, Function, and Bioinformatics*, 71:1175–1182, 2008.

[37] R. Samudrala and J. Moult. An all-atom distance-dependent conditional probability discriminatory function for protein structure prediction. *J. Mol. Biol.*, 275:895–916, 1998.

[38] W.R.P. Scott, P.H. Hünenberger, I.G. Tironi, A.E. Mark, S.R. Billeter, J. Fennen, A.E. Torda, T. Huber, P. Krüger, Gunsteren van, and W.F. The gromos biomolecular simulation program package. *J. Phys. Chem.*, 103:3596–3607, 1999.

[39] Y. Shao and C. Bystroff. Predicting interresidue contacts using templates and pathways. *PROTEINS: Structure, Function, and Bioinformatics*, 53:497–502, 2003.

[40] N. Siew, A. Elofsson, L. Rychlewski, and D. Fischer. Bioinformatics. *Journal Name*, 16(9):776–785, 2000.

[41] K.T. Simons, T. Kooperberg, E. Huang, and D. Baker. Assembly of protein tertiary structures from fragments with similar local sequences using simulated annealing and bayesian scoring functions. *J. Mol. Biol.*, 268:209–225, 1997.

[42] M. Sippl. Recognition of errors in three-dimensional structures of proteins. *PROTEINS: Structure, Function, and Bioinformatics*, 17:355–362, 1993.

[43] A. Sperduti and A. Starita. Supervised neural networks for the classification of structures. *IEEETNN*, 8(3):714–735, 1997.

[44] S.C.E. Tosatto. The victor/frst function for model quality estimation. *J. Comput. Biol.*, 12(10):1316–1327, 2005.

[45] J. Tsai, R. Bonneau, A.V. Morozov, B. Kuhlman, C.A. Rohl, and D. Baker. An improved protein decoy set for testing energy functions for protein structure prediction. *PROTEINS: Structure, Function, and Bioinformatics*, 53:76–87, 2003.

[46] A. Vullo, I. Walsh, and G. Pollastri. A two-stage approach for improved prediction of residue contact maps. *BMC Bioinformatics*, 7:18, 2006.

[47] B. Wallner and A. Elofsson. Can correct protein models be identified? *Protein Science*, 12:1073–1086, 2003.

[48] B. Wallner and A. Elofsson. Identification of correct regions in protein models using structural, alignment, and consensus information. *Protein Science*, 15:900–913, 2006.

[49] B. Wallner and A. Elofsson. Prediction of global and local model quality in casp7 using pcons and proq. *PROTEINS: Structure, Function, and Bioinformatics*, 69(Suppl 8):184–193, 2007.

[50] B. Wallner, H. Fang, and A. Elofsson. Automatic consensus-based fold recognition using pcons, proq, and pmodeller. *PROTEINS: Structure, Function, and Genetics*, 53:534–541, 2003.

[51] S. Wu, J. Skolnick, and Y. Zhang. Ab initio modelling of small proteins by iterative tasser simulations. *BMC Biology*, 5:17, 2007.

[52] Y. Wu, M. Lu, M. Chen, J. Li, and J. Ma. Opus-$c_\alpha$: A knowledge-based potential function requiring only $c_\alpha$ positions. *Protein Science*, 16:1449–1463, 2007.

[53] A. Zemla, C. Venclovas, J. Moult, and K. Fidelis. Processing and analysis of casp3 protein structure predictions. *Proteins*, 37(S3), 1999.

[54] C. Zhang and S.H. Kim. Environment-dependent residue contact energies for proteins. *PNAS*, 97(6):2550–2555, 2000.

[55] Y. Zhang and J. Skolnick. Scoring function for automated assessment of protein structure template quality. *PROTEINS: Structure, Function, and Bioinformatics*, 57:702–710, 2004.

[56] H. Zhou and J. Skolnick. Protein model quality assessment prediction by combining fragment comparisons and a consensus ca contact potential. *PROTEINS: Structure, Function, and Bioinformatics*, 71:1211–1218, 2008.

[57] H. Zhou and Y. Zhou. Distance-scaled, finite ideal-gas reference state improves and stability prediction structure-derived potentials of mean force for structure selection. *Protein Science*, 11:2714–2726, 2002.