

Supplementary Information

SSNN, a method for neural network protein secondary structure fitting using circular dichroism data

5 VINCENT HALL,^{a,b} ANTHONY NASH,^{a,b} ALISON RODGER*^{b,c}

^a Molecular Organisation and Assembly in Cells Doctoral Training Centre, University of Warwick. Coventry, CV4 7AL, UK

^b Department of Chemistry, University of Warwick. Coventry, CV4 7AL, UK. Fax: 44 24 76575795; Tel: 44 24 76574696; E-mail: a.rodger@warwick.ac.uk

^c Warwick Centre for Analytical Science, University of Warwick. Coventry, CV4 7AL, UK.

10

The supplementary information given below includes detailed instructions on how to install SSNN and also the detailed output from SSNNGUI.app and CDsstr for 48 proteins used in leave-one-out mode (SSNN1 and SSNN2 were run for the 48 sets of 47 spectra from CDDATA48).

15 Installing the MATLAB MCR and SSNNGUI

Download the appropriate version of SSNNGUI for your computer from http://www2.warwick.ac.uk/fac/sci/chemistry/research/arodger/arodgergroup/research_intro/instrumentation/ssnn/ (this may take some time).

Unpack or unzip the SSNNGUI package, and use the MCR installer to install the MATLAB Compiler Runtime. Unfortunately, different MCRs are required for different operating systems and the Mathworks MATLAB website only makes available the MCR going back a few versions.

20 Installing the MATLAB Compiler Runtime on a PC should be straightforward if the correct one is chosen; the XP version of SSNN1_2 works fine on Windows 7 Enterprise. If it does not work, make sure you have completed through the MCR installation process.

The situation is more complicated for a Macintosh. For the SSNN_Mac2013a release, one should install MATLAB R2013a MCR, as other MCRs are not compatible. To run this MCR installer launch <InstallForMacOSX.app> (the default will be to install it in the <Applications/MATLAB/MATLAB_Compiler_Runtime> folder). Follow instructions on screen until it asks to you to do something with the

25 DYLD_LIBRARY_PATH and XAPPLRESDIR.

It is *essential* that every character within the relevant <> is correctly typed from the following instructions. Towards the end of the MCR installation, the MCR installer displays messages about DYLD_LIBRARY_PATH and XAPPLRESDIR. The relevant changes can be made in the Terminal application (found within the Utilities folder of <Applications>). In Terminal, type <cd ../> press enter. Repeat typing <cd ../>, pressing enter then typing <ls> enter until the file list includes <Application>. Then type <cd Applications/MATLAB>. Then enter the

30 following (see Figure S1): <export DYLD_LIBRARY_PATH=/Applications/MATLAB/MATLAB_Compiler_Runtime/v81/runtime/maci64:/Applications/MATLAB/MATLAB_Compiler_Runtime/v81/sys/os/maci64:/Applications/MATLAB/MATLAB_Compiler_Runtime/v81/bin/maci64:/System/Library/Frameworks/JavaVM.framework/JavaVM:/System/Library/Frameworks/JavaVM.framework/Libraries>. You should be fine if copying and pasting from the MCR installer window or this paper, but remember to include the <export DYLD_LIBRARY_PATH=> before the /Applications.

35 If an error message *e.g.* <not a valid identifier> appears, check every character in the above text. Then type <export XAPPLRESDIR=/Applications/MATLAB/MATLAB_Compiler_Runtime/v81/X11/app-defaults>, which can also mostly be copied from the MCR installer.

Using the command <env> will show you all of the environment variables, and these should be there now, as above. You can now quit Terminal.

40 Click <next> in the MCR installation window. Click <Finish>. The MCR is now installed.

```
Vince-Hall:MATLAB Vincent$ pwd
/Applications/MATLAB
Vince-Hall:MATLAB Vincent$ export DYLD_LIBRARY_PATH=/Applications/MATLAB/MATLAB_Compiler_Runtime/v81/runtime/maci64:/Applications/MATLAB/MATLAB_Compiler_Runtime/v81/sys/os/maci64:/Applications/MATLAB/MATLAB_Compiler_Runtime/v81/bin/maci64:/System/Library/Frameworks/JavaVM.framework/JavaVM:/System/Library/Frameworks/JavaVM.framework/Libraries
```

Figure S1: The command to export the DYLD library path should look like this in Terminal if your computer is called Vince-Hall.

Using SSNNGUI with pre-trained SSNN1 and SSNN2

5 In the <SSNN_2013a_pkg> folder, launch the <SSNN_2013a.app> application. The best way to do this is using the shell. This is a file that is run in the command line, but does not require the user to input any command line commands. “Right-click” (or ctrl+tap touchpad in Mac) the <run_SSNN_2013a.sh> file, click <Open With Other...>, scroll to Utilities, then <Terminal.app>. The user might have to select <Enable All Applications>. Then click open. A Terminal window should open, and might display an error message. Ignore the error message, just wait a few second, and the SSNN application should launch.

10 Place the required test file ,in the correct format (see <example_3proteins.txt>), in the same folder as <SSNN_2013a.app>. The instructions in the Protocol of Table 1 should then be followed for a protein of known concentration.

To run SSNNGUI first locate the folder with the trained maps and the test spectra by entering the file path in the field marked with <Enter FULL_filepath_from_Users_or_root_no_slash> or click the <Browse> button to find it graphically (select any file within the correct folder). In the next field enter the name of the test file with no file extension (*i.e.* no <.txt> on the end). Enter the number of test spectra in the test file in

15 the appropriate field. Click the “Run SSNN one dataset” button.

SSNNGUI has an option be run with automatic adjustment of concentrations of the test proteins, but that is not covered in this article. So here, ignore the <Run SSNN many concentrations> button, and the <Min concentration>, <Step concentration> and <Max concentration> fields.

The outputs from SSNN3 are the following files:

20 In the top folder

Spectral_NRMSD_vs_concentration_error_protein_1.pdf
best_concentrations.txt
all_NRMSDs.txt
all_NRMSDs_sorted.txt

25

Then in the folders with names such as <53p40x40MAP_5_BMUs_28000ITER_0_06L0_60_Conc>:

Which_are_BMUs1_spectra_of_Winners.txt
Which_are_BMUs2_The_training_set.txt
Which_are_BMUs3_locations_of_Winners.txt

30 *ssnn_residuals.txt*

ssnn_real_spectra.txt
ssnn_predicted_spectra.txt
SSNN_parameters_extended.txt
RMSDspect.txt

35 *plots_with_residuals*

If SSNN is taking too long to run on a laptop computer, it can be put in standby and the run will continue when it is opened again. If there is a problem, open the log viewer (for example Console) to see more output, and any error messages. All of the windows can be closed after the plots of model against experimental CD spectra have been displayed. *Do not close any windows before SSNN has stopped running.* There

40 should be one figure per protein spectrum if running the SSNN one data set. If running the SSNN many concentrations option, the final files are the NRMSD against concentration plots (one per protein spectrum). If you are running the program from the command line (this is the case if running it from the .sh file), then there will be a lot of output from the SSNN.

Running SSNN re-trainable, or “SSNN1_2”

45 Assuming that the appropriate MCR is installed (see Running SSNN3 above), one should be able to run SSNN1_2.app. When the user double-clicks on the SSNN application, it will take a few second to start up (Figure S2).

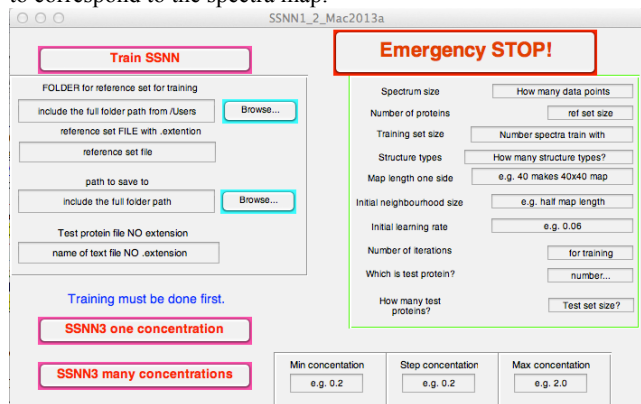
SSNN1

SSNN1 requires a reference set to be loaded in as a .txt file with the following format. All our reference and test data are in this format: the

protein CD spectrum and structure vector written as a column of 57 entries (CD intensities for 240–290 nm at 1 nm spacing followed by 6 structure types summing to 1.00). Other column vector options can be adopted as long as the reference and test spectra are consistent. The column vectors are placed side by side in the same file, in a tab or comma delimited format. For example, the reference set (CDDATA.48+5) used to train this version of SSNN has 53 proteins in an input file of size 57×53. That is CD intensities at 51 nm points, and 6 structure types, for 53 proteins. This input file is loaded using the SSNN graphical user interface (GUI, not “SSNNGUI”), designed for this purpose before SSNN1 is run. The structure information is not used in the training of SSNN1, only the spectral data are used at this stage. SSNN1 outputs a spectra map, which is then used by SSNN2 to give the corresponding structures map, and by SSNN3 to make models of the test protein spectra, and give structure estimations.

10 SSNN2

SSNN2 requires the same input file as SSNN1 as well as the SSNN1 output. The spectra in the input file are used to locate their representations on the spectra map. Then the structures of each protein are added to the structure map at the same coordinates. SSNN2 outputs a structures map to correspond to the spectra map.



15 **Figure S2:** shows the SSNN1_2 application, which incorporates SSNN3.

SSNN1_2.app

The application SSNN1_2.app can be used to run SSNN1, SSNN2, and SSNN3. It is available from

http://www2.warwick.ac.uk/fac/sci/chemistry/research/arodger/arodgergroup/research_intro/instrumentation/ssnn/. The instructions below assume that the pretrained version SSNNGUI has been successfully run by the user. SSNN1_2.app is available in two Macintosh versions: MATLAB R2011a for Mac OS X 10.6.8 and MATLAB R2013a for OS X 10.8, as well as a MATLAB R2011b version for Windows XP, and a R2013a Windows 7 version. This application can be used to train SSNN with any data set of the correct format. We have provided CDDATA.48 and CDDATA.48+5 (see above). For the SSNN1_2_Mac2011a release (see below) the MATLAB R2011a MCR needs to be installed as described above.

25 To run the application, double click on the SSNN1_2*.app icon, then wait for the application to launch. Figure S2 shows the screen which controls the running of the application. SSNN1_2* means some file with a name that starts with “SSNN1_2”.

- i. In the top left there is an editable text field called <FOLDER for reference set for training>. For this field the user has two options: either to type in the full folder path from “/Users/” to the folder in which the reference set file is located or click on “Browse...” to navigate to it graphically. The reference set path and save path do not visibly update, but they are actually entered once the user has clicked on the “Browse...” button..
- ii. The second field is for editable text (<reference set FILE with .extension>), this is the name of the reference set file to train SSNN with. It should be named **with the extension**. It is made to take .txt files. One has to type in the reference set name, there is no option to select it graphically, this is for code reasons.
- iii. The same options are available for the third field as for the first. This is called <path to save to>. This is a folder in which all output .txt files and figures will be saved, it can be the same as the reference path/folder. To reuse a training run of SSNN1_2.app, one needs to change the <Weights and Structure>s folder date to today’s. In such a case they should not worry about entering anything into the second field.
- iv. In the fourth field, <name of text file NO .extension>, the user should type the name of the test spectrum file **without** the .txt **extension**, for folder naming purposes in the code. This file should have number of rows matching “Spectrum size”, and columns matching <How many test proteins>, and contain just spectra in the same range as the reference set file (in our case 51 rows of data from 240 – 190 nm in steps of 1 nm). One has to type in the test set name, there is no option to browse.
- v. On the right side of the GUI, in the box with the thin green border, there are 10 fields. The first is “Spectrum size”, this is where the user should tell SSNN1_2 how many wavelength points there are. (Data are in 1 nm steps in the reference sets provided.)

- vi. The second field on the right is <Number of proteins> to tell SSNN1_2 how many columns there are in the reference set file loaded.
- vii. Field 3 is <Training set size>. This will be smaller than the reference set size only if the user wants to train the SOM in leave-one-out-cross-validation; the SOM can be trained with one protein less than the full set size.
- viii. <Structure types> is the number of structure types that have been assigned in the chosen reference set (6 in those supplied).
- 5 ix. <Map length one side> is so that SSNN can make a square map with the length or width of the number entered here.
- x. < Initial neighbourhood size> is how much of the map should be modified in the first iteration of training to make it more similar to the best matching unit, BMU, at the centre of the neighbourhood and the protein spectrum that was chosen first. A neighbourhood is needed to cluster, otherwise dissimilar spectra would be near each other. The neighbourhood size will decrease with iterations to make the clustering more precise.
- 10 xi. <Initial learning rate> is how fast the BMUs will become more like the reference set. This decreases with iterations.
- xii. < Number of iterations> this is the length of training of the SOM. We found that 28 000 iterations gave the best results (lowest structure estimation NRMSD) with our choice of map size, reference set size *etc.*.
- xiii. <Which is test protein?> allows the user to exclude a protein (referenced by its number describing its position in the list) from the training set. If the user wants to train with the full reference set, then this field should be left empty, and <Training set size> should have the same number as <Number of proteins>.
- 15 xiv. <How many test proteins?> asks how many proteins will be in the test file.
- xv. At the bottom of the GUI there are three fields that are to be used if the intensities of the test spectra need to be scaled, if for example the pathlength or concentration is likely to be in error. The fields are: <Min concentration>, < Step concentration>, and < Max concentration>. These are left empty if not required, they are for use when running “SSNN many concentration”, and are not covered in this article.
- 20

To run the first two modules of SSNN1_2*.app, hit the <Train SSNN> button at the top left. Once the <Train SSNN> button has been clicked, there will be a pause before any output appears, then a window should pop up saying that 100 iterations have been completed, then further windows, counting up in 1000s from 1000. If you bypass training, after modifying the name of the Weights and Structures folder to have today’s date, then the figures will appear much more quickly, see below.

25

SSNN3 part of SSNN1_2

Once the SOM has been trained by running SSNN1 and SSNN2 as outlined above, a <Weights_and_Structures...> folder will have been created. Then the <SSNN3 one concentration> button of SSNN1_2 can be clicked to produce the structure estimations for the test protein(s) as in the SSNNGUI version described above and in the main text. SSNN3 has finished running, when the final protein model against experimental

30 CD spectrum plot pops up. This means one can close all the windows, there should be one for each spectrum in the test set.

For troubleshooting in Mac OS X 10.6.8, Console can be useful, as there will be some rather verbose output from SSNN1_2 appearing here. (This may not work for 10.8.x’s Console, so try running it in the command line using the shell.) The Windows equivalent is Event Viewer. In Linux various applications are available including MultiTail, System Log Viewer, KSystemLog, Xlogmaster and swatch.

Table SII: Structure vectors outputted from SSNN-47 and CDSSTR-47 for all proteins in CDDATA.48 from the CDPro web site tested in a leave-one-out methodology. SSNN was trained with 47 proteins 28,000 iterations, a map size of 40×40, initial neighbourhood of 20, BMUs = 5, $L_0 = 0.1$, $t_1 = 7,000$ iterations, and a $k_1 = 5 \times 10^{-6}$. CDSSTR-47 was run using the executable code available at <http://lamar.colostate.edu/~sreeram/CDPro/main.html>¹ using input files based on 5 CDDATA.48 and SSDATA.48 but with one spectrum removed to be the test spectrum and structure each time. Average SELCON3 performance from² is also included.

Protein number	Protein name	Method	a-regular	a-distorted	b-regular	b-distorted	Tums	Other	Spectral NRMMSD	Structural NRMMSD	Fit minus real: a-regular	Fit minus real: a-distorted	Fit minus real: b-regular	Fit minus real: b-distorted	Fit minus real: tums	Fit minus real: other	CDSSTR NRMMSD
1	a-Bungarotoxin	Real	0.000	0.000	0.014	0.095	0.284	0.608	0.000								
		CDSSTR-47	-0.009	0.032	0.301	0.138	0.209	0.315	0.143	0.530	-0.009	0.032	0.287	0.043	-0.075	-0.293	0.432
		SSNN-52	0.015	0.032	0.086	0.102	0.239	0.526	0.362	0.098	0.015	0.032	0.072	0.007	-0.045	-0.082	
2	Alcohol Dehydrogenase	Real	0.139	0.115	0.139	0.096	0.214	0.297	0.000								
		CDSSTR-47	0.153	0.123	0.169	0.099	0.202	0.248	0.077	0.167	0.014	0.008	0.030	0.003	-0.012	-0.049	0.065
		SSNN-52	0.125	0.104	0.174	0.100	0.220	0.277	0.033	0.102	-0.014	-0.011	0.035	0.004	0.006	-0.020	
3	Adenylate Kinase	Real	0.340	0.206	0.077	0.052	0.012	0.313	0.000								
		CDSSTR-47	0.285	0.164	0.094	0.072	0.187	0.196	0.030	0.428	-0.055	-0.042	0.017	0.020	0.175	-0.117	0.203
		SSNN-52	0.277	0.180	0.074	0.062	0.116	0.291	0.021	0.225	-0.063	-0.026	-0.003	0.010	0.104	-0.022	
4	Azurin	Real	0.047	0.062	0.141	0.109	0.312	0.328	0.000								
		CDSSTR-47	0.118	0.104	0.213	0.121	0.195	0.241	0.115	0.545	0.071	0.042	0.072	0.012	-0.117	-0.087	0.437
		SSNN-52	0.073	0.076	0.170	0.104	0.271	0.307	0.073	0.108	0.026	0.014	0.029	-0.005	-0.041	-0.021	
5	b-lactoglobulin	Real	0.056	0.111	0.287	0.123	0.216	0.207	0.000								
		CDSSTR-47	0.100	0.090	0.155	0.102	0.211	0.337	0.048	0.319	0.044	-0.021	-0.132	-0.021	-0.005	0.130	0.193
		SSNN-52	0.074	0.110	0.245	0.119	0.218	0.233	0.046	0.126	0.018	-0.001	-0.042	-0.004	0.002	0.026	
6	Bence Jones Protein	Real	0.000	0.028	0.294	0.196	0.229	0.252	0.000								
		CDSSTR-47	-0.014	0.004	0.277	0.149	0.227	0.335	0.171	0.118	-0.014	-0.024	-0.017	-0.047	-0.002	0.083	0.032
		SSNN-52	0.003	0.035	0.267	0.172	0.221	0.303	0.263	0.086	0.003	0.007	-0.027	-0.024	-0.008	0.051	
7	Bovine Pancreatic Trypsin	Real	0.069	0.138	0.172	0.069	0.190	0.362	0.000								
		CDSSTR-47	0.122	0.096	0.060	0.044	0.117	0.561	0.114	0.198	0.053	-0.042	-0.112	-0.025	-0.073	0.199	0.114
		SSNN-52	0.074	0.112	0.147	0.071	0.173	0.422	0.137	0.084	0.005	-0.026	-0.025	0.002	-0.017	0.060	
8	Carbonic Anhydrase	Real	0.058	0.104	0.170	0.116	0.240	0.312	0.000								
		CDSSTR-47	0.023	0.034	0.096	0.054	0.108	0.679	0.060	0.255	-0.035	-0.070	-0.074	-0.062	-0.132	0.367	0.019
		SSNN-52	0.043	0.058	0.102	0.068	0.131	0.598	0.054	0.236	-0.015	-0.046	-0.068	-0.048	-0.109	0.286	
9	alpha chymotrypsinogen	Real	0.053	0.082	0.210	0.110	0.210	0.335	0.000								
		CDSSTR-47	0.010	0.035	0.148	0.088	0.176	0.535	0.107	0.173	-0.043	-0.047	-0.062	-0.022	-0.034	0.200	0.003
		SSNN-52	0.050	0.073	0.147	0.078	0.153	0.498	0.068	0.171	-0.003	-0.009	-0.063	-0.032	-0.057	0.163	
10	a-Chymotrypsin	Real	0.069	0.045	0.208	0.106	0.200	0.371	0.000								
		CDSSTR-47	0.028	0.041	0.153	0.073	0.142	0.544	0.062	0.156	-0.041	-0.004	-0.055	-0.033	-0.058	0.173	0.077
		SSNN-52	0.050	0.043	0.175	0.097	0.198	0.437	0.107	0.080	-0.019	-0.002	-0.033	-0.009	-0.002	0.066	
11	Colicin A	Real	0.529	0.225	0.000	0.000	0.044	0.202	0.000								
		CDSSTR-47	0.524	0.280	0.032	0.018	0.051	0.083	0.023	0.110	-0.005	0.055	0.032	0.018	0.007	-0.119	0.061
		SSNN-52	0.499	0.198	0.007	0.012	0.085	0.198	0.024	0.049	-0.030	-0.027	-0.002	0.012	0.041	-0.004	
12	Concanavalin A	Real	0.000	0.038	0.329	0.135	0.236	0.262	0.000								
		CDSSTR-47	0.077	0.076	0.187	0.120	0.205	0.336	0.234	0.290	0.077	0.038	-0.142	-0.015	-0.031	0.074	0.244
		SSNN-52	0.020	0.051	0.324	0.130	0.212	0.263	0.050	0.046	0.020	0.013	-0.005	-0.005	-0.024	0.001	
13	Carboxypeptidase A	Real	0.254	0.127	0.111	0.052	0.212	0.244	0.000								
		CDSSTR-47	0.122	0.108	0.221	0.123	0.223	0.204	0.075	0.680	-0.132	-0.019	0.110	0.071	0.011	-0.040	0.337
		SSNN-52	0.178	0.104	0.199	0.085	0.202	0.232	0.067	0.342	-0.076	-0.023	0.088	0.033	-0.010	-0.012	
14	Cytochrome C	Real	0.214	0.194	0.000	0.000	0.233	0.359	0.000								
		CDSSTR-47	0.204	0.161	0.100	0.076	0.171	0.290	0.053	0.305	-0.010	-0.033	0.100	0.076	-0.062	-0.069	0.145
		SSNN-52	0.198	0.154	0.063	0.056	0.205	0.324	0.054	0.161	-0.016	-0.040	0.063	0.056	-0.028	-0.035	
15	EcoRI Endonuclease	Real	0.192	0.127	0.098	0.080	0.210	0.293	0.000								
		CDSSTR-47	0.209	0.147	0.104	0.078	0.169	0.291	0.049	0.094	0.017	0.020	0.096	-0.002	-0.041	-0.002	0.043
		SSNN-52	0.193	0.146	0.082	0.076	0.216	0.286	0.050	0.052	0.001	0.019	-0.016	-0.004	0.006	-0.007	
16	Elastase	Real	0.021	0.087	0.225	0.117	0.208	0.342	0.000								
		CDSSTR-47	-0.003	0.021	0.159	0.097	0.164	0.551	0.071	0.173	-0.024	-0.066	-0.066	-0.020	-0.044	0.209	0.001
		SSNN-52	0.019	0.057	0.152	0.083	0.155	0.534	0.092	0.172	-0.002	-0.030	-0.073	-0.034	-0.053	0.192	
17	Flavodoxin	Real	0.209	0.108	0.108	0.108	0.264	0.203	0.000								
		CDSSTR-47	0.136	0.110	0.145	0.092	0.210	0.309	0.040	0.273	-0.073	0.002	0.037	-0.016	-0.054	0.106	0.108
		SSNN-52	0.183	0.119	0.095	0.087	0.253	0.263	0.036	0.165	-0.026	0.011	-0.013	-0.021	-0.011	0.060	
18	g-Crystallin	Real	0.006	0.086	0.299	0.161	0.109	0.339	0.000								
		CDSSTR-47	-0.013	0.000	0.235	0.146	0.273	0.340	0.118	0.228	-0.019	-0.086	-0.064	-0.015	0.164	0.001	0.150
		SSNN-52	0.004	0.064	0.306	0.151	0.158	0.317	0.400	0.078	-0.002	-0.022	0.007	-0.010	0.049	-0.022	
19	Green Fluorescent Protein	Real	0.004	0.064	0.347	0.093	0.191	0.301	0.000								
		CDSSTR-47	0.074	0.086	0.219	0.131	0.241	0.249	0.158	0.393	0.070	0.022	-0.128	0.038	0.050	-0.052	0.348
		SSNN-52	0.023	0.058	0.334	0.110	0.192	0.283	0.093	0.045	0.019	-0.006	-0.013	0.017	0.001	-0.018	
20	Glycerinaldehyde-3-phosph	Real	0.172	0.102	0.115	0.093	0.217	0.301	0.000								
		CDSSTR-47	0.178	0.127	0.094	0.071	0.172	0.360	0.029	0.119	0.006	0.025	-0.021	-0.022	-0.045	0.059	0.075
		SSNN-52	0.170	0.119	0.105	0.082	0.223	0.301	0.031	0.044	-0.002	0.017	-0.010	-0.011	0.006	0.000	
21	Glutathione Reductase	Real	0.188	0.142	0.140	0.096	0.172	0.262	0.000								
		CDSSTR-47	0.168	0.127	0.090	0.066	0.167	0.379	0.041	0.174	-0.020	-0.015	-0.050	-0.030	-0.005	0.117	0.029
		SSNN-52	0.148	0.123	0.126	0.096	0.214	0.293	0.034	0.144	-0.040	-0.019	-0.014	0.000	0.042	0.031	
22	Hemoglobin	Real	0.537	0.223	0.000	0.000	0.105	0.136	0.000								
		CDSSTR-47	0.527	0.296	0.039	0.011	0.062	0.062	0.014	0.095	-0.010	0.073	0.039	0.011	-0.043	-0.074	0.064
		SSNN-52	0.563	0.195	0.006	0.005	0.090	0.141	0.045	0.031	0.026	-0.028	0.006	0.005	-0.015	0.005	
23	Hemerythrin	Real	0.478	0.197	0.000	0.000	0.111	0.215	0.000								
		CDSSTR-47	0.416	0.220	0.070	0.060	0.128	0.104	0.020	0.183	-0.062	0.023	0.070	0.060	0.017	-0.111	0.144
		SSNN-52	0.478	0.192	0.024	0.018	0.105	0.184	0.040	0.039	0.000	-0.005	0.024	0.018	-0.006	-0.031	</

Protein number	Protein name	Method	a-regular	a-distorted	b-regular	b-distorted	Turns	Other	Spectral NRMSE	Structural NRMSE	Fit minus real: a-regular	Fit minus real: a-distorted	Fit minus real: b-regular	Fit minus real: b-distorted	Fit minus real: turns	Fit minus real: other	CDSSTR NRMSE	
25	Insulin	Real	0.294	0.235	0.020	0.040	0.050	0.361	0.000									
		CDSSTR-47	0.286	0.236	0.082	0.053	0.191	0.153	0.040	0.454	0.040	-0.008	0.001	0.062	0.013	0.141	-0.208	0.309
		SSNN-52	0.288	0.216	0.026	0.046	0.123	0.300	0.046	0.145	0.046	-0.006	-0.019	0.006	0.006	0.073	-0.061	
26	Lactate Dehydrogenase	Real	0.277	0.161	0.088	0.073	0.155	0.246	0.000									
		CDSSTR-47	0.293	0.176	0.077	0.062	0.165	0.238	0.036	0.060	0.060	0.016	0.015	-0.011	-0.011	0.010	-0.018	-0.028
		SSNN-52	0.290	0.179	0.083	0.067	0.117	0.264	0.025	0.088	0.088	0.013	0.018	-0.005	-0.006	-0.038	0.018	
27	Lysozyme	Real	0.202	0.217	0.016	0.047	0.298	0.221	0.000									
		CDSSTR-47	0.258	0.178	0.066	0.047	0.120	0.326	0.069	0.327	0.066	-0.039	0.050	0.000	0.000	-0.178	0.105	0.179
		SSNN-52	0.195	0.181	0.050	0.063	0.257	0.255	0.050	0.148	0.148	-0.007	-0.039	0.034	0.016	-0.041	0.034	
28	Myoglobin	Real	0.582	0.222	0.000	0.000	0.052	0.144	0.000									
		CDSSTR-47	0.587	0.306	0.021	0.003	0.035	0.049	0.010	0.091	0.091	0.005	0.084	0.021	0.003	-0.017	-0.095	0.031
		SSNN-52	0.659	0.172	0.003	0.002	0.051	0.114	0.031	0.060	0.060	0.077	-0.050	0.003	0.002	-0.001	-0.030	
29	Nuclease	Real	0.094	0.101	0.081	0.107	0.289	0.328	0.000									
		CDSSTR-47	0.182	0.141	0.083	0.058	0.166	0.375	0.047	0.220	0.220	0.088	0.040	0.002	-0.049	-0.123	0.047	0.089
		SSNN-52	0.133	0.121	0.096	0.096	0.249	0.305	0.037	0.131	0.131	0.039	0.020	0.015	-0.011	-0.040	-0.023	
30	Papain	Real	0.059	0.123	0.094	0.075	0.175	0.396	0.000									
		CDSSTR-47	0.040	0.032	0.143	0.108	0.222	0.449	0.079	0.158	0.158	-0.097	-0.091	0.049	0.033	0.047	0.053	0.104
		SSNN-52	0.122	0.112	0.111	0.084	0.195	0.376	0.094	0.055	0.055	-0.015	-0.011	0.017	0.009	0.020	-0.020	
31	Parvalbumin	Real	0.278	0.287	0.000	0.037	0.194	0.204	0.000									
		CDSSTR-47	0.315	0.215	0.067	0.041	0.130	0.232	0.057	0.188	0.188	0.037	-0.072	0.067	0.004	-0.064	0.028	0.069
		SSNN-52	0.268	0.240	0.018	0.038	0.182	0.254	0.078	0.119	0.119	-0.010	-0.047	0.018	0.001	-0.012	0.050	
32	Phosphoglycerate Kinase	Real	0.210	0.135	0.043	0.067	0.231	0.313	0.000									
		CDSSTR-47	0.432	0.217	0.032	0.051	0.110	0.153	0.040	0.317	0.317	0.222	0.082	-0.011	-0.016	-0.121	-0.160	0.186
		SSNN-52	0.259	0.159	0.049	0.058	0.197	0.279	0.056	0.131	0.131	0.049	0.024	0.006	-0.009	-0.034	-0.034	
33	Pepsinogen	Real	0.051	0.154	0.235	0.151	0.165	0.243	0.000									
		CDSSTR-47	0.019	0.041	0.251	0.126	0.230	0.332	0.074	0.214	0.214	-0.032	-0.113	0.016	-0.025	0.065	0.089	0.102
		SSNN-52	0.042	0.109	0.251	0.141	0.193	0.294	0.034	0.111	0.111	-0.009	-0.045	0.016	-0.010	0.028	0.021	
34	Prealbumin	Real	0.031	0.031	0.307	0.142	0.165	0.323	0.000									
		CDSSTR-47	0.013	0.057	0.253	0.116	0.241	0.314	0.075	0.139	0.139	-0.018	0.026	-0.026	0.076	-0.009	0.073	
		SSNN-52	0.036	0.055	0.313	0.136	0.172	0.288	0.037	0.066	0.066	0.005	0.024	0.006	-0.006	0.007	-0.035	
35	Rhodanase	Real	0.150	0.147	0.041	0.068	0.235	0.359	0.000									
		CDSSTR-47	0.218	0.148	0.118	0.080	0.183	0.247	0.018	0.394	0.394	0.068	0.001	0.077	0.012	-0.052	-0.112	0.298
		SSNN-52	0.188	0.151	0.062	0.067	0.207	0.325	0.030	0.096	0.096	0.038	0.001	0.021	-0.001	-0.028	-0.034	
36	Ribonuclease A	Real	0.113	0.097	0.218	0.113	0.218	0.242	0.000									
		CDSSTR-47	0.116	0.092	0.144	0.096	0.220	0.332	0.069	0.201	0.201	0.003	-0.005	-0.074	-0.017	0.002	0.090	0.088
		SSNN-52	0.104	0.093	0.186	0.108	0.228	0.280	0.087	0.112	0.112	-0.009	-0.004	-0.032	-0.005	0.010	0.038	
37	Subtilin BPN	Real	0.171	0.131	0.098	0.080	0.225	0.295	0.000									
		CDSSTR-47	0.042	0.109	0.251	0.141	0.193	0.294	0.034	0.181	0.181	-0.052	-0.042	0.065	0.025	-0.029	0.033	0.151
		SSNN-52	0.166	0.126	0.094	0.078	0.228	0.309	0.055	0.030	0.030	-0.005	-0.006	-0.004	-0.002	0.003	0.014	
38	Subtilin novo	Real	0.113	0.102	0.065	0.073	0.295	0.353	0.000									
		CDSSTR-47	0.211	0.136	0.140	0.097	0.190	0.230	0.046	0.637	0.637	0.098	0.034	0.075	0.024	-0.105	-0.123	0.452
		SSNN-52	0.178	0.155	0.063	0.063	0.223	0.318	0.058	0.185	0.185	0.065	0.053	-0.002	-0.010	-0.072	-0.035	
39	Superoxide Dismutase	Real	0.000	0.018	0.248	0.119	0.298	0.316	0.000									
		CDSSTR-47	0.031	0.047	0.253	0.138	0.214	0.312	0.131	0.140	0.140	0.031	0.029	0.005	0.019	-0.084	-0.004	-0.036
		SSNN-52	0.064	0.066	0.204	0.109	0.236	0.321	0.088	0.176	0.176	0.064	0.048	-0.044	-0.010	-0.062	0.005	
40	T4 Lysozyme	Real	0.421	0.244	0.049	0.037	0.116	0.134	0.000									
		CDSSTR-47	0.269	0.165	0.123	0.079	0.187	0.175	0.026	0.447	0.447	-0.152	-0.079	0.074	0.042	0.071	0.041	0.348
		SSNN-52	0.384	0.216	0.050	0.037	0.109	0.204	0.036	0.099	0.099	-0.037	-0.028	0.001	0.000	-0.007	0.070	
41	Thermolysin	Real	0.282	0.133	0.070	0.095	0.215	0.206	0.000									
		CDSSTR-47	0.258	0.156	0.107	0.065	0.148	0.270	0.028	0.218	0.218	-0.024	0.023	0.037	-0.030	0.064	0.029	
		SSNN-52	0.247	0.166	0.062	0.065	0.178	0.282	0.031	0.189	0.189	-0.035	0.033	-0.008	-0.030	-0.037	0.076	
42	Tumor Necrosis Factor	Real	0.000	0.019	0.293	0.140	0.219	0.329	0.000									
		CDSSTR-47	-0.003	0.019	0.226	0.139	0.221	0.386	0.180	0.092	0.092	-0.003	0.000	-0.067	-0.001	0.002	0.057	0.053
		SSNN-52	0.008	0.038	0.286	0.147	0.208	0.314	0.145	0.040	0.040	0.008	0.019	-0.007	0.007	-0.011	-0.015	
43	Triose Phosphate Isomerase	Real	0.236	0.210	0.090	0.064	0.124	0.276	0.000									
		CDSSTR-47	0.349	0.180	0.075	0.061	0.165	0.175	0.023	0.228	0.228	0.113	-0.030	-0.015	-0.003	0.041	-0.101	0.114
		SSNN-52	0.284	0.198	0.072	0.057	0.142	0.247	0.018	0.114	0.114	0.048	-0.012	-0.018	-0.007	0.018	-0.029	
44	Apo-cytochrome C (5°C)	Real	0.020	0.020	0.020	0.020	0.020	0.900	0.000									
		CDSSTR-47	0.006	0.031	0.082	0.052	0.096	0.725	0.068	0.116	0.116	-0.014	0.011	0.062	0.032	0.076	-0.175	0.106
		SSNN-52	0.020	0.021	0.027	0.023	0.027	0.882	0.017	0.010	0.010	0.000	0.001	0.007	0.003	0.007	-0.018	
45	Apo-cytochrome C (90°C)	Real	0.020	0.020	0.020	0.020	0.020	0.900	0.000									
		CDSSTR-47	0.028	0.034	0.075	0.060	0.113	0.673	0.045	0.163	0.163	0.018	0.014	0.055	0.040	0.093	-0.227	0.014
		SSNN-52	0.030	0.044	0.086	0.055	0.097	0.687	0.068	0.149	0.149	0.010	0.024	0.066	0.035	0.077	-0.213	
46	Ribonuclease (20°C) derivative	Real	0.020	0.020	0.020	0.020	0.020	0.900	0.000									
		CDSSTR-47	0.004	0.036	0.082	0.048	0.101	0.726	0.066	0.116	0.116	-0.016	0.016	0.062	0.028	0.081	-0.174	0.108
		SSNN-52	0.020	0.021	0.024	0.022	0.027	0.886	0.039	0.008	0.008	0.000	0.001	0.004	0.002	0.007	-0.014	
47	Staphylococcal Nuclease	Real	0.020	0.020	0.020	0.020	0.020	0.900	0.000									
		CDSSTR-47	-0.004	0.026	0.054	0.038												