# SUMMER INTERNSHIP REPORT

***on***

## *Classification Of Non-Cancer Proteins from Cancer Proteins Using Random Forest Prediction Model with Amino Acid Sequence Features*

***at***

## National Programme on Technology Enhanced Learning

**Duration: 8 weeks (11th July 2022 to 5th Sept 2022)**

***Submitted to***

**Dr. M. Michael Gromiha**
**Professor**
**Department of Biotechnology**

**Bhupat and Jyoti Mehta School of Biosciences**
**Indian Institute of Technology (IIT) Madras**

*Submitted by*

**Dr. K. Sangeetha**

**Date of Submission: 22/09/2022**

NPTEL

IIT Madras

# Acknowledgement

The internship opportunity I had with **National Programme on Technology Enhanced learning (NPTEL)** in collaboration with **Indian Institute of Technology (IIT) Madras** was a great chance for learning and professional development. Therefore, I consider myself very fortunate to be given this opportunity.

I would like to express greatest gratitude to **Dr. M. Michael Gromiha**, Professor, Department of Biotechnology (IIT-Madras) for heartily welcoming me for the internship and to guide and encourage me through the summer training.

I would also like to thank **Ms. Medha Pandey**, who in spite of her busy doctoral studies, took time to hear, guide and keep me in the right path of concept understanding and data analysis.

I express my deep gratitude to **Mr. Kunaal Naik**, Mentor, Data Science Masterminds to help me understand the science behind model building and make it look easy.

Last but not least, I would like to thank my dear husband**, Mr. Senthil Kumar** for his valuable guidance on programming skills.

# Contents

| Section | Title | Page No. |
|---|---|---|
| 1. | Introduction | 1-3 |
| 2. | Materials and Methods | 3-7 |
| 2.1 | Data Resources | 3 |
| 2.2 | Data pre-processing and derivation of properties | 3-4 |
| 2.3 | Train-test split | 4 |
| 2.4 | Random Forest Classifier | 4 |
| 2.5 | Hyperparameters tuning and feature selection | 5-6 |
| 2.6 | Performance evaluation | 6-7 |
| 3. | Results and discussion | 7-9 |
| 3.1 | Performance of the Random Forest classifier on the test data | 7-9 |
| 4. | Conclusion | 9 |
| 5. | References | 10 |

## List of tables

| Section | Title | Page No. |
|---|---|---|
| 1. | Physiochemical, energetic and conformational acid properties of the 20 amino acids | 11-15 |
| 2. | Distribution of training and test datasets | 4 |
| 3 | Hyperparameters search range for the classifiers | 5 |
| 4 | Hyperparameter tested-value result for the classifiers | 5 |
| 5 | Class-wise top ten important features selected for model preparation | 5-6 |
| 6 | Definitions of performance measures used in the study | 6 |
| 7 | Evaluation of performance of test data | 7 |

## List of figures

| | | |
|---|---|---|
| 1 | Performance of the Random Forest classifier model on the test datasets | 8 |
| 2 | ROC-AUC curves of the Random Forest classifier model on the test datasets | 9 |

# 1. INTRODUCTION:

Researchers' world over has spent over a decade in cancer research and have successfully demonstrated great understanding and advancements in its molecular mechanism. However, in spite of innumerable expert-curations describing the genes/proteins in human system that drive towards cancer, there are several factors having a strong causal impact to cancer that remain functionally uncharacterised. Accurate identification of these genes/proteins is therefore very crucial for investigating into newer targets for cancer therapy.

Probing into the characteristics of proteins, in general, in most mammalian systems, they are mostly found functionally associated with other major building blocks of life such as the nucleic acids, carbohydrates, lipids or even other proteins. Proteins interacting with carbohydrates, known as lectins have shown to aid protein-folding, localization, cell to cell adhesion, proliferation as well as cell death (Gorelik et al., 2001). Recent studies have also evidenced their role in growth and metastasis in malignant cells especially in colon, thyroid and breast cancer (Nangia-Makker et al., 2002) and therefore characterizing these proteins based on their involvement in oncogenesis will help in deeper understanding of cancer causatives.

Other interacting proteins include the nucleic acid binding proteins, the DNA-binding and RNA-binding proteins, known to play vital roles in diverse cellular activities such as the splicing, translation, post-translational modifications are also associated with cancer upon mutation. Structural dysregulation of these genes/proteins result in a multistep process of replicating multiple copies of mutant genes by uncontrolled proliferation of cells (Furney et al., 2006). These cancer-causing genes have been broadly classified as proto-oncogenes and tumor-suppressor genes, depending on gain or loss of function to drive cancer respectively.

Recently, a census of human cancer genes has been compiled in a Cancer Gene Census (CGC) resource which is within the Catalogue of Somatic Mutations in Cancer (COSMIC) describing all those genes with causal effect in human cancer (https://cancer.sanger.ac.uk/census). In spite of the list comprising 719 cancer-causing gene entries, it is evidenced that at least 5-10% genes (i.e., ~25000 putative genes) in the human genome are yet to be identified which could be involved in malignant transformations. By unveiling the common features as well differences between the different group of genes/proteins contributing to cancer state, it is possible to uncover the molecular complexity of cancer as a disease en masse, facilitating the further identification of new drug targets (Furney et al., 2006).

Other biomolecule complexes that are vulnerable to cancerous mutations include the protein-protein interactions (PPI). These PPI contain some amino acids that contribute to the binding more than others and are called hotspots.

In general, the characteristics of any protein or peptide primarily depend on the number and sequence of their amino acids. In recent years, techniques such as Machine learning and deep learning have been widely applied for making predictions on the mutability characteristic of different proteins based on analysis of functional properties such as the composition, compressibility ($K^0$), thermodynamic transfer hydrophobicity ($H_t$), surrounding hydrophobicity index ($H_p$), polarity (P), isoelectric point ($pH_i$), equilibrium constant with reference to the ionization property of COOH group ($pK'$), molecular weight (Mw), bulkiness ($B_1$), chromatographic index ($R_1$), refractive index (µ), normalised consensus hydrophobicity($H_{nc}$), short and medium range non-bonded energy ($E_{sm}$) and more are derived from their primary sequences (Cao et al., 2021).

In this study, we propose a random forest classification strategy to identify gene/proteins are highly likely to be cancer-causing by analysis the cancer properties of carbohydrate, DNA, RNA and protein interacting proteins against the cancer genes curate in CGC, based on the hypothesis that these parameters impact the predisposition of these genes/proteins towards alterations that may lead to a cancer specific phenotype.

## 2. MATERIALS AND METHODS

### *2.1. Data Resources*

The initial datasets for the study contained 1,130 Carbohydrate-binding, 3,859 DNA-binding and 963 RNA-binding protein fasta sequences, 880 protein binding protein heterodimers collected from the UniProt and 768 CGC protein sequences from the CGC resource in fasta format.

### *2.2. Data Pre-processing and derivation of properties*

The duplicate sequences within each of the datasets were first deleted after which 842 unique DNA-binding sequences, 232 RNA-binding, 1130 Carbohydrate-binding and 647 protein heterodimer sequences were remained. The presence of duplicates was also checked in the CGC protein sequences. Each of these protein sequences were then tested for redundancy against the CGC sequences and the matches were filtered out from each of the protein datasets. Further, the sequence identifiers from each dataset were ID-mapped in Uniprot database in order to obtain complete sequence information and each mapped output sequence was again tested for matches with CGC sequences. These matched sequences were also removed from the protein-binding datasets.

About 130 physiochemical, energetic and conformational properties of the 20 amino acids were considered for the study (**Table 1*.**).

The average amino acid compositions for each class of proteins and average values other properties in Table 1 for the 20 amino acids in all the datasets were calculated**.** Further, CD-HIT (v.4.8.1) was used to remove redundant sequences by precluding clusters having identical sequences at different user-set thresholds of 0.7, 0.75, 0.8, 0.85, 0.9 and 0.95. After these removals, the final datasets contained 280 DNA-binding sequences, 163 RNA-binding, 1111 Carbohydrate-binding, 500 protein heterodimer sequences and 693 CGC sequences.

*****Placed in page no.11-15 of the document***

### *2.3. Train-test split*

From the above datasets, 80% of the property records were randomly selected as the training dataset and while the remaining 20% of the features as the independent test datasets (**Table 2.**)

**Table 2:** Distribution of training and test datasets

| Protein Classes | Training | Test |
|---|---|---|
| **DNA-binding** | 778 | 195 |
| **RNA-binding** | 684 | 172 |
| **Protein heterodimer** | 954 | 239 |
| **Carbohydrate-binding** | 1443 | 361 |

### *2.4. Random Forest Classifier*

In this study, we created four classifiers for each class labels of the proteins (carbohydrate, DNA, RNA and protein heterodimer) by applying “one-vs-CGC” binary classification technique using random forest classifier in the scikit-learn package, then trained the model using the training data and finally predict the test data under each class labels.

### *2.5. Hyperparameters tuning and feature selection*

The tuning of hyperparameters (bootstrap, max-depth, max_features, min_samples_leaf, min_samples_split and n_estimators) was done placing appropriate search range (**Table 3**.)

**Table 3:** Hyperparameters search range for the classifiers

| Parameters | Tested value range |
|---|---|
| **Bootstrap** | True, False |
| **max_depth** | 2,4 |
| **Max_features** | Auto, sqrt |
| **Min_samples_split** | 2,5 |
| **n_estimators** | 10, 25, 40, 55, 71, 86, 101, 116,132, 147, 162, 177, 193, 208, 223, 238, 254, 269, 284, 300 |

The best parameters were optimized using grid search **(Table 4.).** The random forest classifier was fitted to compute the feature importance using the sci-kit learn packages.

**Table 4:** Hyperparameter tested-value result for the classifiers

| Protein Class | Parameter tested values | | | | | |
|---|---|---|---|---|---|---|
| | **Bootstrap** | **max_depth** | **max_features** | **min_samples_leaf** | **Min_samples_split** | **n_estimators** |
| **DNA-binding** | True | 4 | auto | 1 | 5 | 10 |
| **RNA-binding** | True | 4 | auto | 1 | 2 | 55 |
| **Protein heterodimer** | True | 4 | auto | 1 | 2 | 116 |
| **Carbohydrate-binding** | False | 4 | auto | 1 | 5 | 40 |

Important features (top 10) were sub-set form the sequences to reduce overfitting problems **(Table 5.).**

**Table 5: Class-wise top ten important features selected for model preparation**

| DNA | | Hetero | | RNA | | Carb | |
|---|---|---|---|---|---|---|---|
| Feature | Importance | Feature | Importance | Feature | Importance | Feature | Importance |
| C | 0.124264 | H | 0.067093 | S | 0.069648161 | Q | 0.123183 |

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| GEIM800108 | 0.05076 | S | 0.047136 | G | 0.048445558 | R | 0.066823 |
| IV0 | 0.050154 | R | 0.041721 | Q | 0.04632393 | P | 0.064244 |
| OOBM850105 | 0.049998 | Q | 0.034027 | C | 0.04547044 | K | 0.062882 |
| RACS770102 | 0.048675 | L | 0.031625 | H | 0.041588343 | S | 0.06186 |
| MAXF760106 | 0.047584 | T | 0.030458 | I | 0.036775609 | E | 0.057489 |
| ASAD | 0.047545 | E | 0.030345 | N | 0.027620529 | H | 0.055449 |
| Hp | 0.044061 | P | 0.03034 | L | 0.026219557 | NADH010101 | 0.048027 |
| Nl | 0.042493 | D | 0.02723 | T | 0.025002648 | C | 0.034995 |
| E | 0.029169 | V | 0.02704 | K | 0.024318277 | ROBB760105 | 0.026609 |

### *2.6. Performance evaluation*

To evaluate performance of prediction, the assessment measurements used include accuracy, recall, precision and F1 score (**Table 6**).

**Table 6**: Definitions of performance measures used in the study

| | |
|---|---|
| **Accuracy** | $\boldsymbol{\frac{TP+TN}{TP+TN+FP+FN}}$ |
| **Recall** | $\frac{TP}{TP+FN}$ |
| **Precision** | $\frac{TP}{TP+FP}$ |
| **F1 score** | $\frac{2*TN}{2*TP+FP+FN}$ |

Where**, TP, TN, FN and FP** are true positives, true negatives, false negatives and false positives, respectively.

The accuracy measures the correctly predicted observations against the total observations. Recall indicates accuracy of the positive samples predicted, while precision is the count of correctly predicted positives against total positive observations and F1 score is the weighted average of precision and recall. For all these measures, 1 is the maximum value indicating best performance that is closeness between predicted and actual observations; 0 is minimum value, which indicates the lowest performance that is the predicted observations differ greatly from the actual observations. Also, an AU-ROC curve (area under the receiver operating characteristics curve) was used to determine the performance of the classification model. An ROC curve features true-positive rates false-positive rates on Y and X-axes respectively, which means greater area under the curve (AUC) is preferred.

## 3. RESULTS AND DISCUSSION

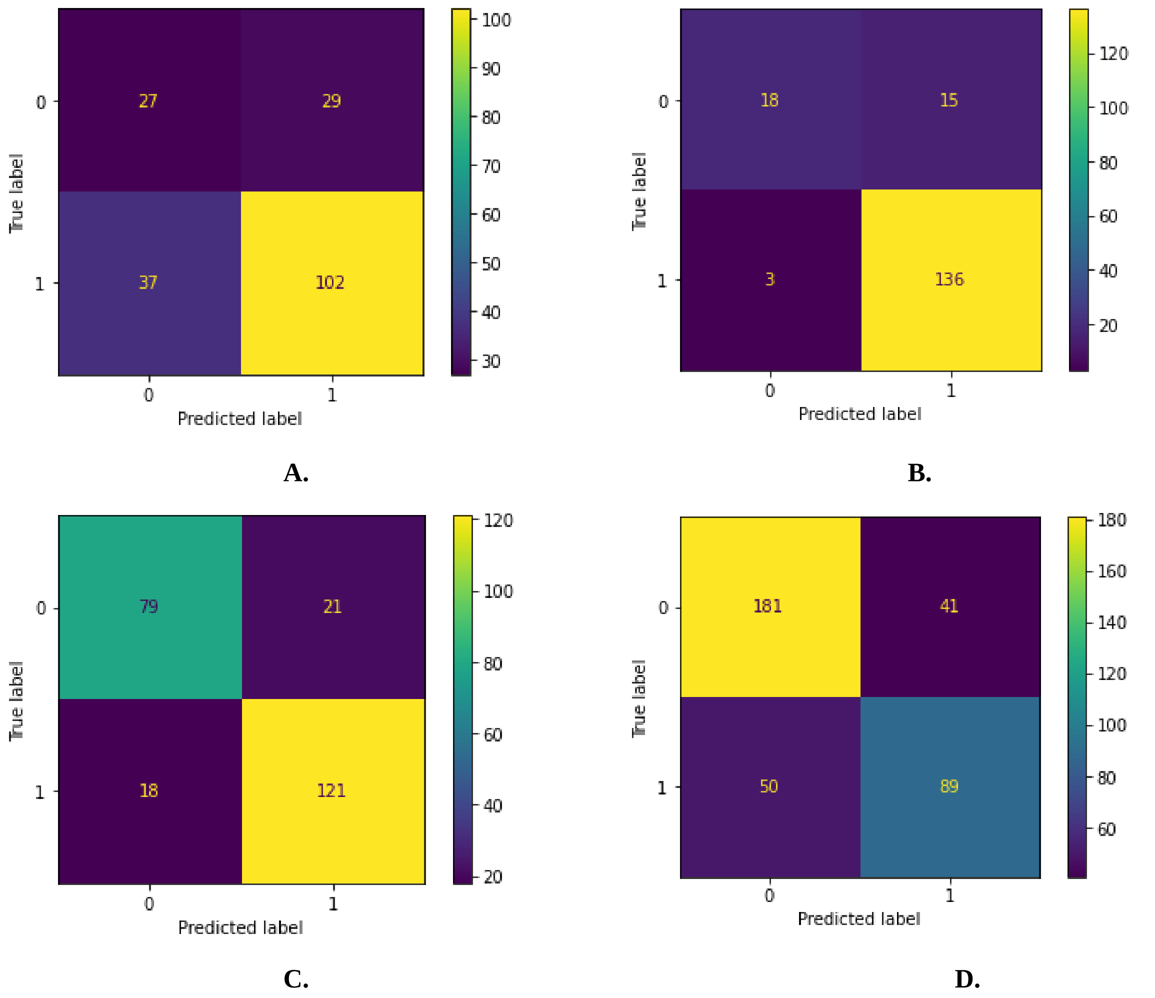
### *3.1. Performance of the Random Forest classifier on the test data*

The performance of Random Forest classifier on the test data for each protein type was evaluated (**Table 7.**).

**Table 7**: Evaluation of performance of test data

| Protein class | Precision | Recall | F1 score |
|---|---|---|---|
| **DNA-binding** | 0.78 | 0.73 | 0.76 |
| **RNA-binding** | 0.90 | 0.98 | 0.94 |
| **Carbohydrate-binding** | 0.68 | 0.64 | 0.66 |
| **Protein-heterodimer** | 0.85 | 0.87 | 0.86 |

The confusion matrix on test data shown in **Fig 1A**. shows the result of random forest classification on the DNA-binding protein test data collected in the study. It shows that 27 of 195 total sequences of DNA-binding protein were correctly identified as DNA sequences and 102 sequences were identified as CGC proteins (cancer genes). Similarly, **Fig1B.** shows 18 of 172 total RNA-binding proteins were rightly identified as RNA and 136 of these as CGC proteins. Confusion matrix in **Fig 1C.** shows 79 of 239 heterodimer proteins to be correctly identified as heterodimers and 121 as CGC and **Fig 1D.** shows 181 of 361 carbohydrate-binding proteins protein were correctly identified as carbohydrate-binding proteins while 89 of these were correctly identified as CGC proteins.



A.

B.

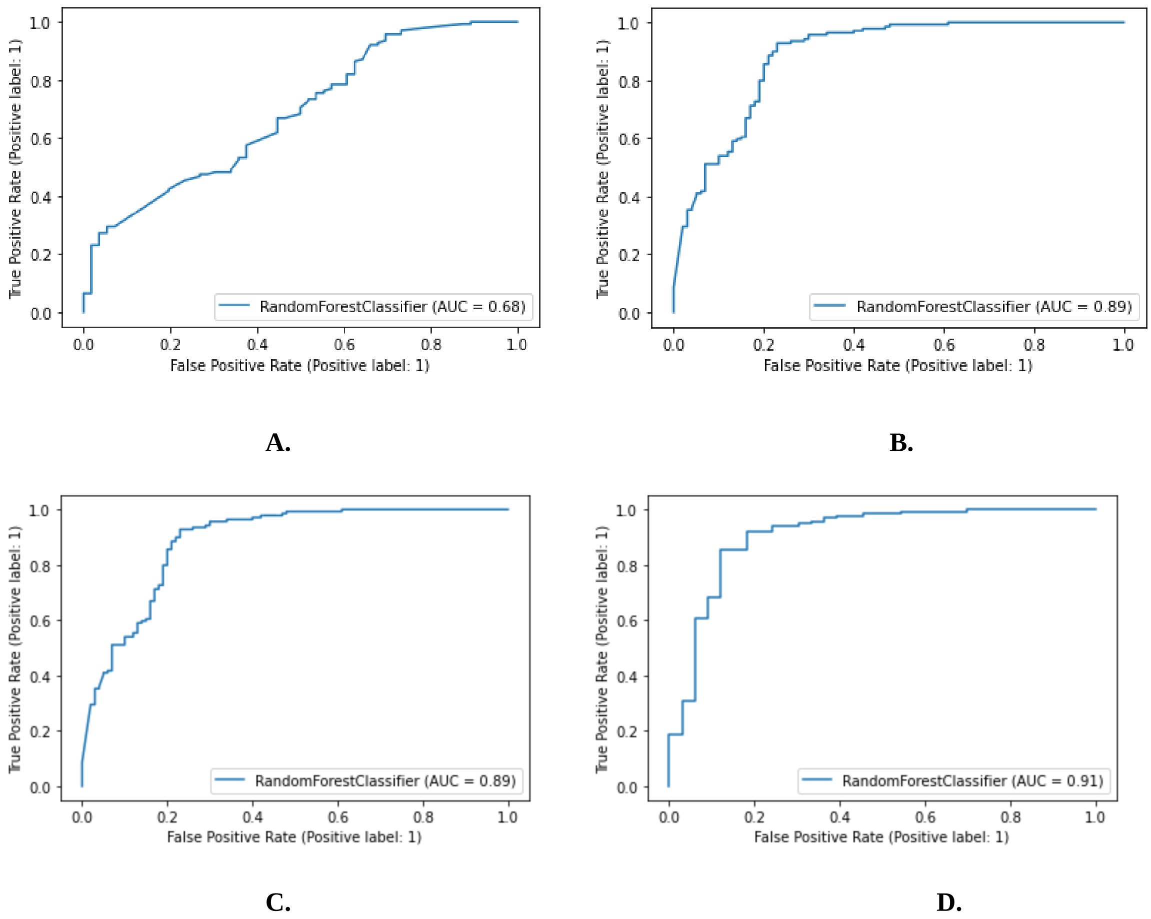
C.

D.

**Fig 1**. Performance of the Random Forest classifier model on the test datasets of **A.** DNA-binding proteins **B**. RNA-binding proteins **C**. heterodimer proteins and **D.** Carbohydrate-binding proteins tested against the sequences in CGC protein dataset.

The evaluation metrics, including precision, recall, F1 scores as mentioned in **Table 6**. and Accuracy (DNA-binding protein: 0.66, RNA-binding protein: 0.90, Heterodimer protein: 0.84, Carbohydrate-binding protein:0.75) to access the performance of the Random Forest classifier indicate the evaluation values of DNA, RNA, Carbohydrate and protein-binding proteins are good, especially the classification of RNA-binding proteins.

The ROC-AUC curves of all the test datasets depict the prediction efficiency for the selected features (**Fig 2A-D**).



A.

B.

C.

D.

**Fig 2**. ROC-AUC curves of the Random Forest classifier model on the test datasets of **A.** DNA-binding proteins **B**. RNA-binding proteins **C**. Heterodimer proteins and **D.** Carbohydrate-binding proteins tested against the sequences in CGC protein dataset.

The confusion matrix, ROC-AUC curves and the evaluations indicate that Random Forest classifier achieves good results on the collected test data.

## 4. CONCLUSION:

This study successfully applies machine learning algorithm to classify cancer genes from the non-cancerous proteins in DNA-binding, RNA-binding, Carbohydrate-binding and protein-binding heterodimer protein datasets. The study can further be extended to classify several other human non-cancer proteins by applying array of machine learning algorithms.

## 5. References

1. Gorelik, E., Galili, U., & Raz, A. (2001) On the role of cell surface carbohydrates and their binding proteins (lectins) in tumor metastasis. *Cancer and Metastasis Reviews*, 20, 245-277.
2. Nangia-Makker, P., Conklin, J., Hogan, V., & Raz, A. (2002) Carbohydrate-binding proteins in cancer, and their ligands as therapeutic agents. *The NDS in Molecular Medicine* 8,187-192.
3. Furney, S.J., Higgins, D.G., Ouzounis, C.A., & Lòpez-Bigas, N. (2006) Structural and functional properties of genes involved in human cancer. *BMC Genomics* 7,3.
4. Cao, Y., Park, S., & Im, W. (2020) A systematic analysis of protein-carbohydrate interactions in the Protein data bank. *Glycobiology* 31, 126-136.
5. Sony, S.M.M., Saraboji, K., Sukumar, N., & Ponnuswamy, M.N. (2006) Role of amino acid properties to determine backbone τ(N-Cα-C') stretching angle in peptides and proteins. *Biophysical Chemistry* 120, 24-31.

**Table 1:** Physiochemical, energetic and conformational acid properties of the 20 amino acids

| Properties | A | D | C | E | F | G | H | I | K | L | M | N | P | Q | R | S | T | V | W | Y |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| K0 | 0.83 | 0.24 | 0.26 | 0 | 0.13 | 0.71 | 0.34 | 0.34 | 0.29 | 0.34 | 0.39 | 0.41 | 1 | 0.28 | 0.74 | 0.49 | 0.38 | 0.43 | 0.46 | 0.09 |
| Ht | 0.23 | 0.18 | 0.4 | 0.18 | 0.76 | 0.03 | 0.23 | 0.84 | 0.44 | 0.58 | 0.44 | 0.02 | 0.73 | 0 | 0.23 | 0.02 | 0.02 | 0.5 | 1 | 0.71 |
| Hp | 0.47 | 0.02 | 0.76 | 0.09 | 0.67 | 0.27 | 0.33 | 1 | 0 | 0.73 | 0.6 | 0.16 | 0.01 | 0.18 | 0.32 | 0.15 | 0.26 | 0.86 | 0.68 | 0.59 |
| P | 0 | 0.96 | 0.03 | 0.96 | 0.01 | 0 | 0.99 | 0 | 0.95 | 0 | 0.03 | 0.07 | 0.03 | 0.07 | 1 | 0.03 | 0.03 | 0 | 0.04 | 0.03 |
| pHi | 0.4 | 0 | 0.29 | 0.31 | 0.34 | 0.4 | 0.6 | 0.41 | 0.87 | 0.4 | 0.37 | 0.33 | 0.44 | 0.36 | 1 | 0.36 | 0.36 | 0.4 | 0.39 | 0.36 |
| pK' | 0.96 | 0.64 | 0.28 | 0.81 | 0.52 | 0.96 | 0.45 | 0 | 0.8 | 0.98 | 0.9 | 0.65 | 0.62 | 0.79 | 0.44 | 0.83 | 0.73 | 0.94 | 1 | 0.82 |
| Mw | 0.11 | 0.45 | 0.36 | 0.56 | 0.7 | 0 | 0.62 | 0.43 | 0.55 | 0.43 | 0.57 | 0.44 | 0.31 | 0.55 | 0.77 | 0.23 | 0.34 | 0.33 | 1 | 0.82 |
| Bl | 0.44 | 0.45 | 0.55 | 0.56 | 0.9 | 0 | 0.56 | 0.99 | 0.68 | 0.99 | 0.71 | 0.52 | 0.77 | 0.61 | 0.6 | 0.33 | 0.68 | 1 | 1 | 0.8 |
| Rf | 0.44 | 0 | 0 | 0.03 | 1 | 0.18 | 0.34 | 0.89 | 0.04 | 0.93 | 0.74 | 0.16 | 0.75 | 0.39 | 0.11 | 0.26 | 0.42 | 0.72 | 0.89 | 0.76 |
| Mu | 0.34 | 0.28 | 0.84 | 0.41 | 0.69 | 0 | 0.51 | 0.45 | 0.5 | 0.44 | 0.51 | 0.31 | 0.26 | 0.41 | 0.63 | 0.15 | 0.26 | 0.33 | 1 | 0.74 |
| Hnc | 0.81 | 0.88 | 0.72 | 0.46 | 0.95 | 0.77 | 0.54 | 1 | 0.26 | 0.92 | 0.81 | 0.45 | 0.68 | 0.43 | 0 | 0.6 | 0.63 | 0.92 | 0.85 | 0.71 |
| Esm | 1 | 0.5 | 0.94 | 0.5 | 0.46 | 0.92 | 0.63 | 0.56 | 0.31 | 0.83 | 0.79 | 0.54 | 0.67 | 0.42 | 0 | 0.79 | 0.69 | 0.69 | 0.23 | 0.23 |
| El | 0.36 | 0.09 | 0.7 | 0.13 | 0.79 | 0.43 | 0.45 | 0.87 | 0 | 0.66 | 0.66 | 0.15 | 0.3 | 0.19 | 0.47 | 0.28 | 0.42 | 0.81 | 1 | 0.66 |
| Et | 0.79 | 0.22 | 1 | 0.25 | 0.73 | 0.79 | 0.58 | 0.87 | 0 | 0.9 | 0.88 | 0.28 | 0.49 | 0.22 | 0.16 | 0.57 | 0.6 | 0.91 | 0.75 | 0.48 |
| Pa | 0.9 | 0.47 | 0.14 | 1 | 0.6 | 0 | 0.46 | 0.54 | 0.63 | 0.68 | 0.94 | 0.11 | 0 | 0.57 | 0.44 | 0.21 | 0.28 | 0.52 | 0.54 | 0.13 |
| Pb | 0.35 | 0.13 | 0.62 | 0 | 0.76 | 0.29 | 0.38 | 0.92 | 0.28 | 0.7 | 0.51 | 0.39 | 0.14 | 0.55 | 0.42 | 0.29 | 0.62 | 1 | 0.75 | 0.83 |
| Pt | 0.17 | 0.91 | 0.66 | 0.25 | 0.12 | 1 | 0.44 | 0 | 0.5 | 0.11 | 0.12 | 1 | 0.96 | 0.47 | 0.44 | 0.88 | 0.45 | 0.03 | 0.45 | 0.61 |
| Pc | 0.12 | 0.61 | 0.59 | 0.25 | 0.12 | 0.91 | 0.47 | 0.07 | 0.39 | 0.1 | 0 | 0.76 | 1 | 0.27 | 0.47 | 0.74 | 0.48 | 0.04 | 0.17 | 0.47 |
| Ca | 0.15 | 0.27 | 0.25 | 0.42 | 0.69 | 0 | 0.5 | 0.54 | 0.69 | 0.46 | 0.62 | 0.31 | 0.19 | 0.48 | 0.88 | 0.15 | 0.31 | 0.42 | 1 | 0.69 |
| F | 0.57 | 0.96 | 0.38 | 0.81 | 0 | 1 | 0.23 | 0.15 | 0.96 | 0.21 | 0.19 | 0.74 | 1 | 0.81 | 0.77 | 0.94 | 0.57 | 0.21 | 0.17 | 0.68 |
| Br | 0.64 | 0.19 | 1 | 0.09 | 0.89 | 0.64 | 0.51 | 0.98 | 0 | 0.87 | 0.72 | 0.21 | 0.26 | 0.13 | 0.06 | 0.36 | 0.36 | 0.83 | 0.68 | 0.42 |
| Ra | 0.32 | 0.14 | 0.21 | 0.26 | 0.82 | 0.23 | 0.3 | 1 | 0 | 0.69 | 0.58 | 0.06 | 0.06 | 0.15 | 0.13 | 0.11 | 0.14 | 0.91 | 0.76 | 0.21 |
| Ns | 0.39 | 0.02 | 1 | 0.07 | 0.58 | 0.43 | 0.31 | 0.88 | 0 | 0.84 | 0.51 | 0.05 | 0.26 | 0.19 | 0.28 | 0.22 | 0.31 | 0.92 | 0.7 | 0.62 |
| aN | 0.83 | 0.28 | 0.17 | 0.76 | 0.6 | 0.28 | 0.47 | 0.64 | 0.59 | 1 | 0.65 | 0.28 | 0 | 0.51 | 0.35 | 0.37 | 0.39 | 0.74 | 0.7 | 0.3 |
| aC | 0.58 | 0.97 | 0.21 | 0.9 | 0.34 | 0.13 | 0.09 | 0.16 | 0.11 | 0.11 | 0.19 | 0.3 | 1 | 0.45 | 0 | 0.23 | 0.48 | 0.13 | 0.56 | 0.18 |
| aM | 0.55 | 0.25 | 0.69 | 0.57 | 0.51 | 0.18 | 1 | 0.35 | 0.74 | 0.47 | 0.66 | 0.42 | 0 | 0.73 | 0.71 | 0.39 | 0.21 | 0.54 | 0.21 | 0.23 |
| V0 | 0.17 | 0.3 | 0.24 | 0.42 | 0.78 | 0 | 0.55 | 0.64 | 0.65 | 0.64 | 0.62 | 0.35 | 0.39 | 0.5 | 0.84 | 0.17 | 0.33 | 0.47 | 1 | 0.8 |
| Nm | 0.83 | 0.51 | 0.59 | 0.81 | 0.69 | 0.22 | 0.69 | 0.47 | 0.67 | 0.92 | 1 | 0.55 | 0 | 0.75 | 0.65 | 0.26 | 0.26 | 0.33 | 0.61 | 0.37 |

| | | | | | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Nl | 0.42 | 0.05 | 0.99 | 0 | 0.63 | 0.56 | 0.37 | 1 | 0.02 | 0.65 | 0.5 | 0.32 | 0.3 | 0.12 | 0.37 | 0.36 | 0.48 | 0.95 | 0.74 | 0.77 |
| Hgm | 0.6 | 0.06 | 0.97 | 0 | 0.62 | 0.48 | 0.6 | 0.95 | 0.05 | 0.67 | 0.6 | 0.4 | 0.24 | 0.3 | 0.42 | 0.49 | 0.32 | 0.78 | 1 | 0.61 |
| ASAD | 0.22 | 0.43 | 0.43 | 0.65 | 0.79 | 0 | 0.68 | 0.72 | 0.89 | 0.64 | 0.85 | 0.45 | 0.45 | 0.67 | 1 | 0.28 | 0.42 | 0.52 | 0.99 | 0.9 |
| ASAN | 0.17 | 0.5 | 0 | 0.7 | 0.17 | 0.13 | 0.44 | 0.12 | 1 | 0.15 | 0.26 | 0.48 | 0.48 | 0.6 | 0.85 | 0.34 | 0.38 | 0.11 | 0.24 | 0.36 |
| dASA | 0.22 | 0.21 | 0.57 | 0.29 | 0.84 | 0 | 0.52 | 0.8 | 0.35 | 0.69 | 0.84 | 0.24 | 0.24 | 0.4 | 0.58 | 0.15 | 0.27 | 0.58 | 1 | 0.82 |
| dGh | 0.86 | 0.48 | 0.69 | 0.36 | 0.78 | 0.86 | 0.41 | 1 | 0.6 | 0.99 | 0.85 | 0.38 | 1 | 0.32 | 0 | 0.34 | 0.64 | 0.97 | 0.34 | 0.05 |
| GhD | 0.91 | 0.5 | 0.81 | 0.41 | 0.86 | 0.9 | 0.54 | 0.99 | 0.51 | 0.98 | 0.9 | 0.46 | 1 | 0.42 | 0 | 0.49 | 0.68 | 0.97 | 0.6 | 0.32 |
| GhN | 0.96 | 0.53 | 0.93 | 0.46 | 0.93 | 0.94 | 0.66 | 0.98 | 0.73 | 0.98 | 0.95 | 0.54 | 1 | 0.52 | 0 | 0.63 | 0.73 | 0.97 | 0.84 | 0.57 |
| dHh | 0.92 | 0.67 | 0.79 | 0.55 | 0.6 | 1 | 0.42 | 0.74 | 0.61 | 0.78 | 0.71 | 0.54 | 0.95 | 0.48 | 0.03 | 0.51 | 0.68 | 0.82 | 0.18 | 0 |
| -TdSh | 0.19 | 0.16 | 0.21 | 0.24 | 0.74 | 0 | 0.6 | 0.76 | 0.45 | 0.68 | 0.62 | 0.29 | 0.32 | 0.33 | 0.83 | 0.29 | 0.36 | 0.56 | 1 | 0.96 |
| dCph | 0.29 | 0 | 0.17 | 0.01 | 0.93 | 0.05 | 0.44 | 1 | 0.38 | 0.91 | 0.74 | 0.03 | 0.53 | 0.03 | 0.35 | 0.09 | 0.34 | 0.76 | 0.89 | 0.71 |
| dGc | 0.13 | 0.47 | 0.45 | 0.57 | 0.52 | 0.16 | 0.62 | 0 | 0.33 | 0.01 | 0.22 | 0.53 | 0 | 0.59 | 0.81 | 0.55 | 0.31 | 0.03 | 0.86 | 1 |
| dHc | 0.12 | 0.26 | 0.56 | 0.34 | 0.81 | 0 | 0.49 | 0.21 | 0.18 | 0.19 | 0.44 | 0.18 | 0.06 | 0.25 | 0.36 | 0.35 | 0.24 | 0.17 | 0.93 | 1 |
| -TdSc | 0.78 | 0.83 | 0.34 | 0.79 | 0 | 0.98 | 0.6 | 0.52 | 0.84 | 0.56 | 0.33 | 1 | 0.76 | 0.95 | 0.95 | 0.76 | 0.72 | 0.61 | 0.1 | 0.09 |
| dG | 0.24 | 0.22 | 0.62 | 0.2 | 1 | 0.28 | 0.44 | 0.22 | 0.14 | 0.22 | 0.43 | 0.14 | 0.23 | 0.17 | 0 | 0.11 | 0.16 | 0.23 | 0.87 | 0.56 |
| dH | 0.44 | 0.41 | 0.86 | 0.4 | 1 | 0.37 | 0.46 | 0.41 | 0.26 | 0.41 | 0.65 | 0.21 | 0.39 | 0.24 | 0 | 0.38 | 0.4 | 0.42 | 0.79 | 0.72 |
| -TdS | 0.49 | 0.53 | 0.06 | 0.54 | 0 | 0.6 | 0.53 | 0.53 | 0.69 | 0.53 | 0.28 | 0.77 | 0.55 | 0.74 | 1 | 0.53 | 0.52 | 0.51 | 0.24 | 0.22 |
| v | 0.1 | 0.4 | 0.2 | 0.5 | 0.7 | 0 | 0.6 | 0.4 | 0.5 | 0.4 | 0.4 | 0.4 | 0.3 | 0.5 | 0.7 | 0.2 | 0.3 | 0.3 | 1 | 0.8 |
| s | 0 | 0.4 | 0 | 0.6 | 0.4 | 0 | 0.4 | 0.2 | 0 | 0.4 | 0 | 0.4 | 0 | 0.6 | 1 | 0 | 0.2 | 0.2 | 0.4 | 0.4 |
| f | 0 | 0.4 | 0.2 | 0.6 | 0.4 | 0 | 0.4 | 0.4 | 0.8 | 0.4 | 0.6 | 0.4 | 0 | 0.6 | 1 | 0.2 | 0.2 | 0.2 | 0.4 | 0.4 |
| Pf-s | 0.12 | 0.5 | 0.18 | 0.15 | 0.1 | 0.68 | 0.53 | 0.01 | 0.32 | 0.07 | 0.1 | 1 | 0 | 0.27 | 0.28 | 0.21 | 0.01 | 0.02 | 0.12 | 0.18 |
| GEIM800105 | 0.84 | 1.27 | 0.59 | 0.57 | 1.15 | 0.94 | 0.81 | 1.29 | 0.86 | 1.1 | 0.88 | 0.66 | 0.8 | 1.02 | 1.04 | 1.05 | 1.2 | 1.56 | 1.15 | 1.39 |
| GEIM800108 | 0.91 | 0.93 | 1.4 | 0.97 | 0.72 | 1.51 | 0.9 | 0.65 | 0.82 | 0.59 | 0.58 | 1.64 | 1.66 | 0.94 | 1 | 1.23 | 1.04 | 0.6 | 0.67 | 0.92 |
| BIOV880102 | 44 | 90 | -91 | -139 | 148 | -8 | 47 | 100 | -188 | 108 | 121 | -72 | -36 | -117 | -68 | -60 | -54 | 117 | 163 | 22 |
| GRAR740102 | 8.1 | 5.5 | 13 | 12.3 | 5.2 | 9 | 10.4 | 5.2 | 11.3 | 4.9 | 5.7 | 11.6 | 8 | 10.5 | 10.5 | 9.2 | 8.6 | 5.9 | 5.4 | 6.2 |
| GRAR740103 | 31 | 55 | 54 | 83 | 132 | 3 | 96 | 111 | 119 | 111 | 105 | 56 | 32.5 | 85 | 124 | 32 | 61 | 84 | 170 | 136 |
| HOPA770101 | 1 | 0.1 | 6.5 | 6.2 | 1.4 | 1.1 | 2.8 | 0.8 | 5.3 | 0.8 | 0.7 | 2.2 | 0.9 | 2.1 | 2.3 | 1.7 | 1.5 | 0.9 | 1.9 | 2.1 |
| ISOY800102 | 0.86 | 1.39 | 0.69 | 0.66 | 1.16 | 0.7 | 1.06 | 1.31 | 0.77 | 1.01 | 1.06 | 0.74 | 1.16 | 0.89 | 0.98 | 1.09 | 1.24 | 1.4 | 1.17 | 1.28 |
| ISOY800103 | 0.78 | 0.6 | 1.5 | 0.97 | 0.67 | 1.73 | 0.83 | 0.4 | 1.01 | 0.57 | 0.3 | 1.56 | 1.55 | 0.78 | 1.06 | 1.19 | 1.09 | 0.44 | 0.74 | 1.14 |
| ISOY800104 | 1.09 | 0.5 | 0.77 | 0.92 | 0.5 | 1.25 | 0.67 | 0.66 | 1.25 | 0.44 | 0.45 | 1.14 | 2.96 | 0.83 | 0.97 | 1.21 | 1.33 | 0.56 | 0.62 | 0.94 |

| | | | | | | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| JOND750101 | 0.87 | 1.52 | 0.66 | 0.67 | 2.87 | 0.1 | 0.87 | 3.15 | 1.64 | 2.17 | 1.67 | 0.09 | 2.77 | 0 | 0.85 | 0.07 | 0.07 | 1.87 | 3.77 | 2.67 |
| JUKT750101 | 5.3 | 1.3 | 3.6 | 3.3 | 2.3 | 4.8 | 1.4 | 3.1 | 4.1 | 4.7 | 1.1 | 3 | 2.5 | 2.4 | 2.6 | 4.5 | 3.7 | 4.2 | 0.8 | 2.3 |
| KANM800102 | 0.81 | 1.17 | 0.71 | 0.53 | 1.2 | 0.88 | 0.92 | 1.48 | 0.77 | 1.24 | 1.05 | 0.62 | 0.61 | 0.98 | 0.85 | 0.92 | 1.18 | 1.66 | 1.18 | 1.23 |
| KANM800103 | 1.45 | 0.7 | 0.91 | 1.29 | 1.2 | 0.53 | 1.13 | 1.23 | 1.27 | 1.56 | 1.83 | 0.64 | 0.21 | 1.14 | 1.15 | 0.48 | 0.77 | 1.1 | 1.17 | 0.74 |
| KARP850101 | 1.041 | 0.96 | 1.033 | 1.094 | 0.93 | 1.142 | 0.982 | 1.002 | 1.093 | 0.967 | 0.947 | 1.117 | 1.055 | 1.165 | 1.038 | 1.169 | 1.073 | 0.982 | 0.925 | 0.961 |
| KRIW710101 | 4.6 | -1 | 5.7 | 5.6 | 3.2 | 7.6 | 4.5 | 2.6 | 7.9 | 3.25 | 1.4 | 5.9 | 7 | 6.1 | 6.5 | 5.25 | 4.8 | 3.4 | 4 | 4.35 |
| KRIW790101 | 4.32 | 1.73 | 6.04 | 6.17 | 2.59 | 6.09 | 5.66 | 2.31 | 7.92 | 3.93 | 2.44 | 6.24 | 7.19 | 6.13 | 6.55 | 5.37 | 5.16 | 3.31 | 2.78 | 3.58 |
| LAWE840101 | -0.48 | -0.32 | -0.75 | -0.71 | 1.03 | 0 | -0.51 | 0.81 | -0.09 | 1.02 | 0.81 | -0.87 | 2.03 | -0.32 | -0.06 | 0.05 | -0.35 | 0.56 | 0.66 | 1.24 |
| LEVM760101 | -0.5 | -1 | 2.5 | 2.5 | -2.5 | 0 | -0.5 | -1.8 | 3 | -1.8 | -1.3 | 0.2 | -1.4 | 0.2 | 3 | 0.3 | -0.4 | -1.5 | -3.4 | -2.3 |
| LIFS790101 | 0.92 | 1.16 | 0.48 | 0.61 | 1.25 | 0.61 | 0.93 | 1.81 | 0.7 | 1.3 | 1.19 | 0.6 | 0.4 | 0.95 | 0.93 | 0.82 | 1.12 | 1.81 | 1.54 | 1.53 |
| LIFS790103 | 0.9 | 1.24 | 0.47 | 0.62 | 1.23 | 0.56 | 1.12 | 1.54 | 0.74 | 1.26 | 1.09 | 0.62 | 0.42 | 1.18 | 1.02 | 0.87 | 1.3 | 1.53 | 1.75 | 1.68 |
| MANP780101 | 12.97 | 14.63 | 10.85 | 11.89 | 14 | 12.43 | 12.16 | 15.67 | 11.36 | 14.9 | 14.39 | 11.42 | 11.37 | 11.76 | 11.72 | 11.23 | 11.69 | 15.71 | 13.93 | 13.42 |
| MAXF760102 | 0.86 | 1.17 | 0.72 | 0.62 | 1.16 | 0.97 | 1.06 | 1.24 | 0.79 | 0.98 | 1.08 | 0.74 | 1.22 | 0.89 | 0.94 | 1.04 | 1.18 | 1.33 | 1.07 | 1.25 |
| MAXF760106 | 1 | 1.09 | 1.39 | 1.04 | 0.65 | 0.46 | 0.71 | 0.68 | 1.05 | 1.01 | 0.36 | 0.87 | 1.95 | 1.13 | 1.18 | 1.56 | 1.23 | 0.58 | 1.1 | 0.87 |
| MIYS850101 | 2.36 | 3.36 | 1.67 | 1.74 | 4.37 | 2.06 | 2.41 | 4.17 | 1.23 | 3.93 | 4.22 | 1.7 | 1.89 | 1.75 | 1.92 | 1.81 | 2.04 | 3.49 | 3.82 | 2.91 |
| NAGK730102 | 0.96 | 1.13 | 0.9 | 0.33 | 1.37 | 0.9 | 0.87 | 1.54 | 0.81 | 1.26 | 1.29 | 0.72 | 0.75 | 1.18 | 0.67 | 0.77 | 1.23 | 1.41 | 1.13 | 1.07 |
| NAGK730103 | 0.72 | 1.01 | 1.04 | 0.75 | 0.58 | 1.35 | 0.76 | 0.8 | 0.84 | 0.63 | 0.62 | 1.38 | 1.43 | 0.81 | 1.33 | 1.34 | 1.03 | 0.83 | 0.87 | 1.35 |
| BURA740102 | 0.288 | 0.533 | 0.271 | 0.262 | 0.318 | 0.312 | 0.2 | 0.411 | 0.265 | 0.4 | 0.375 | 0.229 | 0.34 | 0.327 | 0.362 | 0.354 | 0.388 | 0.495 | 0.231 | 0.429 |
| ARGP820101 | 0.61 | 1.07 | 0.46 | 0.47 | 2.02 | 0.07 | 0.61 | 2.22 | 1.15 | 1.53 | 1.18 | 0.06 | 1.95 | 0 | 0.6 | 0.05 | 0.05 | 1.32 | 2.65 | 1.88 |
| NISK860101 | -0.22 | 4.66 | -4.12 | -3.64 | 5.27 | -1.62 | 1.28 | 5.58 | -4.18 | 5.01 | 3.51 | -2.65 | -3.03 | -2.76 | -0.93 | -2.84 | -1.2 | 4.45 | 5.2 | 2.15 |
| OOBM770103 | -0.491 | -0.67 | -0.356 | -0.371 | -0.729 | -0.534 | -0.54 | -0.762 | -0.3 | -0.65 | -0.659 | -0.382 | -0.463 | -0.405 | -0.554 | -0.455 | -0.515 | -0.728 | -0.839 | -0.656 |
| OOBM850103 | 0.46 | 0.2 | -0.33 | 0.48 | 0.52 | 0.64 | -1.31 | 3.28 | -1.71 | 0.43 | 0.15 | 1.31 | -0.58 | -1.12 | -1.54 | -0.83 | -1.52 | 0.54 | 1.25 | -2.21 |
| CHAM820101 | 0.046 | 0.128 | 0.105 | 0.151 | 0.29 | 0 | 0.23 | 0.186 | 0.219 | 0.186 | 0.221 | 0.134 | 0.131 | 0.18 | 0.291 | 0.062 | 0.108 | 0.14 | 0.409 | 0.298 |
| OOBM850105 | 4.55 | -0.78 | 2.85 | 5.16 | 4.37 | 9.14 | 4.48 | 2.1 | 10.68 | 3.24 | 2.18 | 5.56 | 5.14 | 4.15 | 5.97 | 6.78 | 8.6 | 3.81 | 1.97 | 2.4 |
| PONP800102 | 7.62 | 10.93 | 6.18 | 6.38 | 8.99 | 7.31 | 7.85 | 9.99 | 5.72 | 9.37 | 9.83 | 6.17 | 6.64 | 6.67 | 6.81 | 6.93 | 7.08 | 10.38 | 8.41 | 8.53 |
| PONP800103 | 2.63 | 3.36 | 2.29 | 2.31 | 3.02 | 2.55 | 2.57 | 3.08 | 2.12 | 2.98 | 3.18 | 2.27 | 2.46 | 2.45 | 2.45 | 2.6 | 2.55 | 3.21 | 2.85 | 2.79 |
| PONP800107 | 3.7 | 3.03 | 2.6 | 3.3 | 6.6 | 3.13 | 3.57 | 7.69 | 1.79 | 5.88 | 5.21 | 2.12 | 2.12 | 2.7 | 2.53 | 2.43 | 2.6 | 7.14 | 6.25 | 3.03 |
| PRAM900104 | 0.78 | 0.8 | 1.41 | 1 | 0.58 | 1.64 | 0.69 | 0.51 | 0.96 | 0.59 | 0.39 | 1.28 | 1.91 | 0.97 | 0.88 | 1.33 | 1.03 | 0.47 | 0.75 | 1.05 |
| RACS770101 | 0.934 | 0.9 | 0.994 | 0.986 | 0.773 | 1.015 | 0.882 | 0.766 | 1.04 | 0.825 | 0.804 | 0.986 | 1.047 | 1.047 | 0.962 | 1.056 | 1.008 | 0.825 | 0.848 | 0.931 |

| | | | | | | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| RACS770102 | 0.941 | 0.866 | 1.071 | 1.1 | 0.723 | 1.055 | 0.911 | 0.742 | 1.232 | 0.798 | 0.781 | 1.038 | 1.093 | 1.15 | 1.112 | 1.082 | 1.043 | 0.817 | 0.867 | 1.05 |
| RACS820113 | 17.05 | 28.84 | 19.27 | 20.12 | 16.26 | 38.14 | 23.07 | 16.66 | 16.46 | 10.89 | 20.61 | 34.81 | 23.94 | 15.42 | 21.25 | 19.95 | 18.92 | 17.06 | 23.36 | 26.49 |
| CHOC760101 | 115 | 135 | 150 | 190 | 210 | 75 | 195 | 175 | 200 | 170 | 185 | 160 | 145 | 180 | 225 | 115 | 140 | 155 | 255 | 230 |
| ROBB760105 | -2.3 | 4.4 | -4.4 | -5 | 2.6 | -4.2 | -2.5 | 6.7 | -3.3 | 2.3 | 2.3 | -4.1 | -1.8 | 1.2 | 0.4 | -1.7 | 1.3 | 6.8 | -1 | 4 |
| ROSM880102 | -0.67 | -2 | 1.57 | 1.78 | -3.24 | 0 | 1.09 | -3.02 | 2.46 | -3.02 | -1.67 | 2.27 | -1.75 | 2.12 | 3.89 | 0.1 | -0.42 | -2.18 | -2.86 | 0.98 |
| SIMZ760101 | 0.73 | 0.7 | 0.54 | 0.55 | 2.65 | 0 | 1.1 | 2.97 | 1.5 | 2.49 | 1.3 | -0.01 | 2.6 | -0.1 | 0.73 | 0.04 | 0.44 | 1.69 | 3 | 2.97 |
| WOEC730101 | 7 | 5.5 | 13 | 12.5 | 5 | 7.9 | 8.4 | 4.9 | 10.1 | 4.9 | 5.3 | 10 | 6.6 | 8.6 | 9.1 | 7.5 | 6.6 | 5.6 | 5.3 | 5.7 |
| CHOP780202 | 0.83 | 1.19 | 0.54 | 0.37 | 1.38 | 0.75 | 0.87 | 1.6 | 0.74 | 1.3 | 1.05 | 0.89 | 0.55 | 1.1 | 0.93 | 0.75 | 1.19 | 1.7 | 1.37 | 1.47 |
| YUTK870102 | 6.8 | 8.3 | 7 | 4.9 | 8.3 | 6.4 | 9.2 | 10 | 7.5 | 12.2 | 8.4 | 6.2 | 6.9 | 8.5 | 0 | 8 | 7 | 9.4 | 5.7 | 6.8 |
| ZIMJ680101 | 0.83 | 1.48 | 0.64 | 0.65 | 2.75 | 0.1 | 1.1 | 3.07 | 1.6 | 2.52 | 1.4 | 0.09 | 2.7 | 0 | 0.83 | 0.14 | 0.54 | 1.79 | 0.31 | 2.97 |
| ZIMJ680102 | 11.5 | 13.46 | 11.68 | 13.57 | 19.8 | 3.4 | 13.69 | 21.4 | 15.71 | 21.4 | 16.25 | 12.82 | 17.43 | 14.45 | 14.28 | 9.47 | 15.77 | 21.57 | 21.67 | 18.03 |
| ZIMJ680105 | 9.9 | 2.8 | 2.8 | 3.2 | 18.8 | 5.6 | 8.2 | 17.1 | 3.5 | 17.6 | 14.9 | 5.4 | 14.8 | 9 | 4.6 | 6.9 | 9.5 | 14.3 | 17.1 | 15 |
| CHOP780204 | 1.29 | 0.66 | 2.02 | 2.44 | 0.61 | 0.76 | 0.73 | 0.67 | 0.66 | 0.58 | 0.71 | 0.81 | 2.01 | 1.22 | 0.44 | 0.74 | 1.08 | 0.61 | 1.47 | 0.68 |
| ONEK900102 | -0.77 | -0.23 | -0.15 | -0.27 | -0.41 | 0 | -0.06 | -0.23 | -0.65 | -0.62 | -0.5 | -0.07 | 3 | -0.33 | -0.68 | -0.35 | -0.11 | -0.14 | -0.45 | -0.17 |
| VINM940101 | 0.984 | 0.906 | 1.068 | 1.094 | 0.915 | 1.031 | 0.95 | 0.927 | 1.102 | 0.935 | 0.952 | 1.048 | 1.049 | 1.037 | 1.008 | 1.046 | 0.997 | 0.931 | 0.904 | 0.929 |
| NADH010101 | 58 | 116 | -97 | -131 | 92 | -11 | -73 | 107 | -24 | 95 | 78 | -93 | -79 | -139 | -184 | -34 | -7 | 100 | 59 | -11 |
| NADH010102 | 51 | 137 | -78 | -115 | 108 | -13 | -55 | 106 | -205 | 103 | 73 | -84 | -79 | -128 | -144 | -26 | -3 | 108 | 69 | 11 |
| NADH010104 | 32 | 182 | -29 | -74 | 132 | -22 | -25 | 106 | -124 | 104 | 82 | -73 | -82 | -95 | -95 | -34 | 20 | 113 | 118 | 44 |
| NADH010105 | 24 | 194 | 0 | -57 | 131 | -28 | -31 | 102 | -9 | 103 | 90 | -76 | -85 | -87 | -79 | -36 | 34 | 111 | 116 | 43 |
| FUKS010102 | 6.77 | 0.31 | 8.57 | 12.93 | 1.92 | 7.95 | 2.8 | 2.72 | 10.2 | 4.43 | 1.87 | 5.5 | 4.79 | 5.24 | 6.87 | 5.41 | 5.36 | 3.57 | 0.54 | 2.26 |
| FUKS010103 | 7.43 | 0.42 | 8.71 | 5.86 | 1.18 | 9.4 | 1.49 | 1.76 | 9.67 | 2.74 | 0.6 | 9.12 | 5.6 | 5.42 | 4.51 | 9.6 | 8.95 | 3.1 | 1.18 | 3.26 |
| KUHL950101 | 0.78 | 0.55 | 1.35 | 1.45 | 0.47 | 0.68 | 0.99 | 0.47 | 1.1 | 0.56 | 0.66 | 1.2 | 0.69 | 1.19 | 1.58 | 1 | 1.05 | 0.51 | 0.7 | 1 |
| ZHOH040101 | 2.18 | 3.89 | 1.75 | 1.89 | 5.88 | 1.17 | 2.51 | 4.5 | 2.12 | 4.71 | 3.63 | 1.85 | 2.09 | 2.16 | 2.71 | 1.66 | 2.18 | 3.77 | 6.46 | 5.01 |
| ZHOH040102 | 1.79 | 2.22 | 2.33 | 2.52 | 4.84 | 0.7 | 3.06 | 4.59 | 2.5 | 4.72 | 3.91 | 2.83 | 2.45 | 2.37 | 3.2 | 1.82 | 2.45 | 3.67 | 5.64 | 4.46 |
| ZHOH040103 | 13.4 | 22.6 | 8.2 | 7.3 | 23.9 | 7 | 11.3 | 20.3 | 6.1 | 20.8 | 15.7 | 7.6 | 9.9 | 8.5 | 8.5 | 8.2 | 10.3 | 19.5 | 24.5 | 19.5 |
| PONJ960101 | 91.5 | 114.4 | 135.2 | 154.6 | 198.8 | 67.5 | 163.2 | 162.6 | 162.5 | 163.4 | 165.9 | 138.3 | 123.4 | 156.4 | 196.1 | 102 | 126 | 138.4 | 209.8 | 237.2 |
| WOLR790101 | 1.12 | 0.59 | -0.83 | -0.92 | 0.67 | 1.2 | -0.93 | 1.16 | -0.8 | 1.18 | 0.55 | -0.83 | 0.54 | -0.78 | -2.55 | -0.05 | -0.02 | 1.13 | -0.19 | -0.23 |
| OLSK800101 | 1.38 | 1.43 | 0.52 | 0.71 | 1.72 | 1.34 | 0.66 | 2.32 | 0.15 | 1.47 | 1.78 | 0.37 | 0.85 | 0.22 | 0 | 0.86 | 0.89 | 1.99 | 0.82 | 0.47 |
| KIDA850101 | -0.27 | -1.05 | 0.81 | 1.17 | -1.43 | -0.16 | 0.28 | -0.77 | 1.7 | -1.1 | -0.73 | 0.81 | -0.75 | 1.1 | 1.87 | 0.42 | 0.63 | -0.4 | -1.57 | -0.56 |

| | | | | | | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| CORJ870102 | -0.414 | 0.162 | -1.31 | -1.218 | 1.938 | -0.684 | -0.63 | 1.237 | -0.67 | 1.215 | 1.02 | -0.916 | -0.503 | -0.905 | -0.584 | -0.563 | -0.289 | 0.899 | 0.514 | 1.699 |
| CORJ870104 | -0.26 | 0.83 | -1.3 | -0.73 | 1.09 | -0.4 | -0.18 | 1.1 | -1.01 | 1.52 | 1.09 | -0.46 | -0.62 | -0.83 | 0.08 | -0.55 | -0.71 | 1.15 | -0.13 | 0.69 |
| CORJ870106 | -1.35 | 4.37 | -11.88 | -4.56 | 11.35 | -5.82 | 6.54 | 10.93 | -11.92 | 9.88 | 7.47 | -10.96 | -10.86 | -1.34 | -3.89 | -6.21 | -4.83 | 8.2 | 1.8 | 7.61 |
| CORJ870108 | 1.37 | -4.47 | 8.93 | 4.04 | -7.96 | 3.39 | -1.65 | -7.92 | 7.7 | -8.68 | -7.13 | 6.29 | 6.25 | 3.88 | 1.33 | 4.08 | 4.02 | -6.94 | 0.79 | -4.73 |
| MIYS990101 | -0.02 | -0.96 | 0.72 | 0.74 | -2.22 | 0.38 | 0 | -1.89 | 1.01 | -2.29 | -1.36 | 0.63 | 0.47 | 0.56 | 0.44 | 0.55 | 0.25 | -1.34 | -1.28 | -0.88 |
| FASG890101 | -0.21 | -6.04 | 1.36 | 2.3 | -4.65 | 0 | -1.23 | -4.81 | 3.88 | -4.68 | -3.66 | 0.96 | 0.75 | 1.52 | 2.11 | 1.74 | 0.78 | -3.5 | -3.32 | -1.01 |
| DAYM780201 | 100 | 20 | 106 | 102 | 41 | 49 | 66 | 96 | 56 | 40 | 94 | 134 | 56 | 93 | 65 | 120 | 97 | 74 | 18 | 41 |
| EISD860101 | 0.67 | 0.38 | -1.2 | -0.76 | 2.3 | 0 | 0.64 | 1.9 | -0.57 | 1.9 | 2.4 | -0.6 | 1.2 | -0.22 | -2.1 | 0.01 | 0.52 | 1.5 | 2.6 | 1.6 |
| FASG760101 | 89.09 | 121.15 | 133.1 | 147.13 | 165.19 | 75.07 | 155.16 | 131.17 | 146.19 | 131.17 | 149.21 | 132.12 | 115.13 | 146.15 | 174.2 | 105.09 | 119.12 | 117.15 | 204.24 | 181.19 |
| FASG760102 | 297 | 178 | 270 | 249 | 284 | 290 | 277 | 284 | 224 | 337 | 283 | 236 | 222 | 185 | 238 | 228 | 253 | 293 | 282 | 344 |
| FAUJ830101 | 0.31 | 1.54 | -0.77 | -0.64 | 1.79 | 0 | 0.13 | 1.8 | -0.99 | 1.7 | 1.23 | -0.6 | 0.72 | -0.22 | -1.01 | -0.04 | 0.26 | 1.22 | 2.25 | 0.96 |
| FAUJ880101 | 1.28 | 1.77 | 1.6 | 1.56 | 2.94 | 0 | 2.99 | 4.19 | 1.89 | 2.59 | 2.35 | 1.6 | 2.67 | 1.56 | 2.34 | 1.31 | 3.03 | 3.67 | 3.21 | 2.94 |
| FAUJ880106 | 2.04 | 3.41 | 3.78 | 3.31 | 6.02 | 1 | 5.66 | 3.49 | 4.87 | 4.45 | 4.8 | 4.37 | 4.31 | 3.53 | 6.24 | 2.7 | 3.17 | 3.17 | 5.9 | 6.72 |
| BIGC670101 | 52.6 | 68.3 | 68.4 | 84.7 | 113.9 | 36.3 | 91.9 | 102 | 105.1 | 102 | 97.7 | 75.7 | 73.6 | 89.7 | 109.1 | 54.9 | 71.2 | 85.1 | 135.4 | 116.2 |

Abbreviations: $K_0$ = compressibility; $H_t$ = thermodynamic transfer hydrophobicity; $H_p$ = surrounding hydrophobicity; P = polarity; $pH_i$ = isoelectric point; pKV = equilibrium constant with reference to the ionization property of COOH group; Mw = molecular weight; $B_1$ = bulkiness; $R_l$ = chromatographic index; l = refractive index; $H_{nc}$ = normalized consensus hydrophobicity; $E_{sm}$ = short and medium range non-bonded energy; $E_l$ = long range non-bonded energy; $E_t$ = total non-bonded energy ($E_{sm}$+$E_l$); $P_a$, $P_h$, $P_t$ and $P_c$ = respectively, a-helical, h-structure, turn and coil tendencies; $C_a$ = helical contact area; F = mean rms fluctuational displacement; $B_r$ = buriedness; $R_a$ = solvent accessible reduction ratio; $N_s$ = average number of surrounding residues; $a_n$, $a_c$ and $a_m$ = respectively, power to be at the N-terminal, C-terminal and middle of a-helix.;$V_0$ = partial-specific volume; $N_m$ and $N_l$ = respectively, average medium and long-range contacts; $H_{gm}$ = combined surrounding hydrophobicity (globular and membrane); $ASA_D$, $ASA_N$ and DASA = respectively, solvent accessible surface area for denatured, native and unfolding; $DG_h$, $G_{hD}$ and $G_{hN}$ = Respectively, Gibbs free energy change of hydration for unfolding, denatured and native protein; $DH_h$ = unfolding enthalpy change of hydration; _$TDS_h$ = unfolding entropy change of hydration; $DC_{ph}$ = Unfolding hydration heat capacity change; $DG_c$, $DH_c$ and _$TDS_c$ = respectively, unfolding Gibbs free energy, unfolding enthalpy and unfolding entropy changes of chain; DG, DH, and _TDS = respectively, unfolding Gibbs free energy change, unfolding enthalpy change and unfolding entropy change; v = volume (number of non-hydrogen side chain atoms); s = shape (position of branch point in a side-chain); f = flexibility (number of side-chain dihedral angles). $K_0$ in $m_3$/mol/Pa (_10_$_{15}$); $H_t$, $H_p$, $H_{nc}$, $H_{gm}$, $DG_h$, $G_{hD}$, $G_{hN}$, $DH_h$, _$TDS_h$, $DG_c$, $DH_c$, _$TDS_c$, DG, DH and _TDS in kcal/mol; P in Debye; $pH_i$ and pkV in pH units; $E_{sm}$, $E_l$ and $E_t$ in kcal/mol/atom; $B_1$, $C_a$, $ASA_D$, $ASA_N$ and DASA in A° $_2$; F in A°; $V_0$ in $m_3$/mol (_10_$_6$); $DC_{ph}$ in cal/mol/K and the rest are dimensionless quantities (Sony et al., 2006);

Amino acid abbreviations: A = Alanine, D: Aspartic acid, E: Glutamic acid, F: Phenylalanine, G: Glycine, H: Histidine, I: Isoleucine, K: Lysine, M: Methionine, N: Asparagine, P: Proline, Q: Glutamine, R: Arginine, S: Serine, T: Threonine, V: Valine, W: Tryptophan, Y: Tyrosine