Supplementary Information for: Machine Learning Based Predictive Modeling of Glass Transition Temperatures: A Case of Polyhydroxyalkanoate Homo- and Co-polymers

Ghanshyam Pilania*

Materials Science and Technology Division,

Los Alamos National Laboratory, Los Alamos, NM 87545, USA

Carl N. Iverson

Chemistry Division, Los Alamos National Laboratory, Los Alamos, NM 87545, USA

Turab Lookman

Theoretical Division, Los Alamos National Laboratory, Los Alamos, NM 87545, USA

Babetta L. Marrone

Bioscience Division, Los Alamos National Laboratory, Los Alamos, NM 87545, USA

Table S I: **The polymer dataset employed for the machine learning model development for T_g predictions.**

Monomer A	Monomer B	Monomer A ratio	Copolymer nature	Mol. weight (in 10^3)	PDI	T_g (K)	Ref.
3H5PhP		100.0	R	350.0	3.50	287.15	[1]
3H5PhP	3H7PhHp	23.0	R	156.0	2.33	261.95	[2]
3H5PhP		100.0	R	79.0	2.20	288.85	[3]
3H6PhH		100.0	R	475.0	2.20	271.85	[2]
3H6PhH	3H4PhB	89.4	R	113.0	2.70	278.45	[3]
3H6PhH	3H8PhO	27.0	R	165.0	2.01	258.35	[2]
3HB		100.0	R	1623.0	1.22	276.15	[4]
3HB		100.0	R	166.0	2.60	277.15	[5]
3HB		100.0	R	3240.0	1.70	281.15	[6]
3HB	2HB	91.0	R	100.0	3.10	272.05	[7]
3HB	2HP	97.5	R	309.0	2.71	273.15	[7]
3HB	2HP	96.3	R	303.0	2.68	273.15	[7]
3HB	2HP	96.4	R	423.0	2.73	273.15	[7]
3HB	2HP	97.9	R	278.0	2.36	273.15	[7]
3HB	3H4MeV	60.4	B	1050.0	2.50	269.15	[8]
3HB	3H4MeV	67.7	B	1250.0	2.70	269.15	[8]
3HB	3H4MeV	73.8	B	1340.0	3.10	271.15	[8]
3HB	3HHx	89.0		346.0	1.22	271.57	[9]
3HB	3HP	24.0		652.4	2.80	262.15	[10]
3HB	3HP	33.0		704.0	3.20	263.15	[11]
3HB	3HP	29.0		704.0	3.20	262.15	[11]
3HB	3HP	54.0		684.6	6.52	273.15	[10]
3HB	3HP	22.0		420.0	3.00	259.15	[11]
3HB	3HP	12.0		319.0	2.90	258.15	[11]
3HB	3HP	75.0		208.1	2.19	273.15	[10]

Continued on next page

Table S I – *Continued from previous page*

Monomer A	Monomer B	Monomer A ratio	Copolymer nature	Mol. weight (in 10^3)	PDI	T_g (K)	Ref.
3HB	3HP	80.0		588.0	2.10	272.15	[11]
3HB	3HP	89.0		735.0	2.10	274.15	[11]
3HB	3HP	57.0		672.0	2.40	268.15	[11]
3HB	3HP	63.0		638.0	2.20	268.15	[11]
3HB	3HP	93.0		704.0	2.20	276.15	[11]
3HB	3HP	71.0		609.0	2.90	271.15	[11]
3HB	3HP	85.0		382.0	2.29	277.15	[10]
3HB	3HV	55.0		960.0	2.40	263.15	[12]
3HB	3HV	54.0		450.0	2.30	267.40	[13]
3HB	3HV	36.0		3526.4	3.20	260.15	[14]
3HB	3HV	27.0		3324.6	3.11	262.15	[14]
3HB	3HV	29.0		508.0	2.00	260.15	[12]
3HB	3HV	32.0		410.0	2.00	266.55	[13]
3HB	3HV	68.0	R	550.0	1.40	266.55	[15]
3HB	3HV	35.0		2774.4	2.78	269.15	[14]
3HB	3HV	44.0		3452.7	3.08	272.15	[14]
3HB	3HV	12.0		4252.8	3.20	259.15	[14]
3HB	3HV	12.0		3857.2	3.03	259.15	[14]
3HB	3HV	8.0		4997.2	2.48	258.15	[14]
3HB	3HV	37.0		3286.1	2.88	272.15	[14]
3HB	3HV	13.1		3580.0	2.45	263.25	[16]
3HB	3HV	8.0		5240.4	2.64	258.15	[14]
3HB	3HV	6.0		4853.3	2.31	258.15	[14]
3HB	3HV	24.0		2930.4	2.91	270.15	[14]
3HB	3HV	5.0		4849.7	2.42	258.15	[14]
3HB	3HV	65.0	R	290.0	5.60	263.45	[15]

Continued on next page

Table S I – *Continued from previous page*

Monomer A	Monomer B	Monomer A ratio	Copolymer nature	Mol. weight (in 10^3)	PDI	T_g (K)	Ref.
3HB	3HV	70.0	R	250.0	2.90	269.25	[15]
3HB	3HV	0.5		1690.0	2.17	255.28	[16]
3HB	3HV	71.0	R	410.0	1.40	269.45	[15]
3HB	3HV	37.0	R	1700.0	2.20	267.15	[6]
3HB	3HV	76.0		452.2	1.90	267.15	[12]
3HB	3HV	69.9	R	1730.0	1.43	270.04	[17]
3HB	3HV	22.0	R	1830.0	1.90	271.15	[6]
3HB	3HV	90.9	R	1560.0	1.41	275.40	[17]
3HB	3HV	88.0		2542.0	2.48	276.15	[14]
3HB	3HV	84.0		2608.6	2.56	276.15	[14]
3HB	3HV	87.0	R	1800.0	2.61	262.15	[18]
3HB	3HV	97.5	R	1120.0	1.07	276.55	[19]
3HB	3HV	100.0		2965.0	2.84	276.15	[14]
3HB	3HV	100.0		3620.0	3.06	276.15	[14]
3HB	3HV	100.0		3581.8	3.12	276.15	[14]
3HB	3HV	100.0		3173.1	2.96	276.15	[14]
3HB	3HV	97.4	R	1732.0	1.20	275.35	[19]
3HB	3HV	79.4		3400.0	2.48	272.65	[16]
3HB	3HV	57.0		3909.8	2.98	275.15	[14]
3HB	3HV	62.0		3733.9	2.89	274.15	[14]
3HB	3HV	28.5		3070.0	2.44	263.55	[16]
3HB	3HV	55.6		2710.0	1.76	267.65	[16]
3HB	3HV	66.7	B	2000.0	1.63	264.25	[17]
3HB	3HV	53.2	B	1470.0	1.31	265.85	[17]
3HB	4HB	17.0	R	126.0	2.80	233.15	[5]
3HB	4HB	24.0	R	126.0	4.10	236.15	[5]

Continued on next page

Table S I – *Continued from previous page*

Monomer A	Monomer B	Monomer A ratio	Copolymer nature	Mol. weight (in 10^3)	PDI	T_g (K)	Ref.
3HB	4HB	60.0	R	793.0	2.50	257.15	[20]
3HB	4HB	9.0	R	60.0	1.54	220.15	[21]
3HB	4HB	46.0	R	165.0	3.10	251.15	[5]
3HB	4HB	11.0	R	61.0	1.49	218.15	[21]
3HB	4HB	66.0	R	150.0	2.80	270.15	[5]
3HB	4HB	15.0	R	87.0	2.23	232.15	[21]
3HB	4HB	35.0	R	47.0	1.82	242.15	[22]
3HB	4HB	17.0	R	98.0	2.09	232.15	[21]
3HB	4HB	70.0	R	376.0	3.55	256.15	[22]
3HB	4HB	80.0	R	391.0	3.60	257.15	[21]
3HB	4HB	79.0	R	375.0	3.10	257.15	[22]
3HB	4HB	71.0	R	110.0	5.00	270.15	[5]
3HB	4HB	90.0	R	342.0	2.63	261.15	[22]
3HB	4HB	89.6		570.0	3.80	276.55	[23]
3HB	4HB	89.0		320.0	2.67	269.15	[24]
3HB	4HB	94.2	R	1800.0	3.27	255.15	[18]
3HB	4HB	87.0	R	153.0	2.80	272.15	[5]
3HB	4HB	95.7	R	1300.0	3.10	273.15	[18]
3HB	4HB	87.0	R	74.0	5.00	269.15	[5]
3HB	4HB	85.0		750.0	3.00	270.55	[23]
3HB	4HB	92.0	R	88.0	6.20	272.15	[5]
3HB	4HB	90.6		380.0	1.81	269.15	[24]
3HB	4HB	94.8	R	1700.0	3.15	253.15	[18]
3HB	4HB	95.3		580.0	2.52	273.15	[24]
3HB	4HB	97.0	R	58.0	6.90	274.15	[5]
3HB	4HB	100.0	R	2200.0	3.44	277.15	[18]

Continued on next page

Table S I – *Continued from previous page*

Monomer A	Monomer B	Monomer A ratio	Copolymer nature	Mol. weight (in 10^3)	PDI	T_g (K)	Ref.
3HB	4HB	100.0		640.0	2.56	276.15	[24]
3HB	4HB	19.7	B	50.0	3.10	225.85	[25]
3HB	4HB	65.0	R	351.0	6.20	253.15	[21]
3HB	4HB	61.0	R	330.0	5.19	250.15	[22]
3HB	4HB	53.0	R	51.0	1.74	243.15	[22]
3HB		100.0		1635.0	1.93	270.07	[9]
3HB		100.0	R	153.0	3.30	281.05	[3]
3HB		100.0	R	505.0	1.87	275.65	[22]
3HB		100.0		226.0	1.13	275.65	[26]
3HB		100.0	R	2200.0	3.44	277.15	[18]
3HB		100.0		1459.2	1.90	277.15	[11]
3HD		100.0	R	361.0	1.45	235.94	[27]
3HH		100.0	R	272.0	1.93	244.96	[27]
3HHp		100.0		455.0	1.81	241.02	[9]
3HHx		100.0		272.0	1.32	244.96	[9]
3HO	3HB	98.0		180.0	1.22	234.77	[9]
3HP	4HB	18.2		302.0	1.28	243.67	[28]
3HP	4HB	88.1		284.0	1.29	248.73	[28]
3HP	4HB	74.5		259.0	1.32	241.87	[28]
3HP	4HB	62.1		280.0	1.30	237.01	[28]
3HP	4HB	33.0		278.0	1.31	231.28	[28]
3HP	4HB	63.0	B	209.0	2.62	239.77	[29]
3HP	4HB	71.0	B	217.0	2.52	239.75	[29]
3HP	4HB	75.0		172.0	2.83	241.56	[29]
3HP		100.0		200.2	1.40	254.15	[11]
3HP		100.0		163.0	1.48	255.30	[28]

Continued on next page

Table S I – *Continued from previous page*

Monomer A	Monomer B	Monomer A ratio	Copolymer nature	Mol. weight (in 10^3)	PDI	T_g (K)	Ref.
3HV		100.0		230.0	3.50	257.35	[26]
3HV		100.0		1056.0	1.30	258.06	[9]
4HB		100.0		854.0	1.75	227.48	[9]
4HB		100.0		389.0	1.17	226.15	[28]

Table S II: A description of QSPR-based features used in this study. Specific sources used for different features are also explicitly identified.

Name	Symbol	Ref.	Description
Chi0, Chi1	Cn (n = 1–2)	[30]	Simple molecular connectivity index
Chi0v–Chi4v	Cn_v (n = 1–4)	[30]	Valence molecular connectivity indices
BalabanJ	B_j	[31]	A chemical distance-based topological index
Molecular weight	M_{Wt}	[32]	Monomer's molecular weight
LabuteASA	L_{ASA}	[33]	Labute's approximate surface area
Molar refractivity	Mol_R	[34]	Molar refractivity
Molar volume	Mol_V	[35]	Molar volume
SMR_VSA1–SMR_VSA10	VSA_n (n = 1–10)	[36]	Molecular operating environment (MOE)-type descriptors using molar refractivity contributions and surface area contributions
Maximum absolute partial charge	MPC	[37]	Maximum absolute partial Gasteiger atomic charge
Eccentricity	Ecc	[38]	Molecular eccentricity from handbook of chemoinformatics
Asphericity	Asp	[38]	Molecular asphericity from from handbook of chemoinformatics
Inertial shape factor	ISF	[38]	Inertial shape factor based on molecule's principal moment of inertia
Normalized principal moments ratios	NPR_1, NPR_2	[39]	Normalized principal moments ratios
Radius of gyration	R_G	[40]	Arteca's radius of gyration
Spherocity index	SI	[38]	Spherocity index from handbook of chemoinformatics
FractionCSP3	F_{Csp3}	[37]	Fraction of sp^3 hybridized carbons
Normalized number of rotatable bonds	NRB/M_{Wt}	[37]	Number of rotatable bonds normalized by the molecular weight

SISSO Rank	Descriptor Formula	Correlation with T_g
1	(((FCSP3/Ecc)*(MPC/BalabanJ))/((Ecc/NRBwt)+(VSA5/Chi2v)))	0.947
2	((FCSP3/(Ecc*BalabanJ))/((Ecc/NRBwt)+(VSA5/Chi2v)))	0.946
3	(((FCSP3+NRBwt)/(Ecc*BalabanJ))/((Ecc/NRBwt)+(VSA5/Chi2v)))	0.946
4	(((Chi2v/VSA5)+(FCSP3/BalabanJ))/((Ecc-NRBwt)/cbrt(NRBwt)))	0.946
5	(((NRBwt/ROG)+(FCSP3/BalabanJ))/((Ecc/NRBwt)+(VSA5/Chi2v)))	0.946
6	(((NRBwt/VSA5)*(VSA5+Chi2v))*((FCSP3/ROG)+(Ecc/BalabanJ)))	0.946
7	(((SI/VSA5)+(FCSP3/BalabanJ))/((Ecc/NRBwt)+(VSA5/Chi2v)))	0.946
8	(((Chi2v/VSA5)+(FCSP3/BalabanJ))/((Ecc*MPC)/cbrt(NRBwt)))	0.946
9	(((FCSP3/ROG)+(Ecc/BalabanJ))/((Chi2v-VSA5)/(VSA5*NRBwt)))	0.946
10	(((FCSP3/ROG)+(Ecc/BalabanJ))*((MPC+BalabanJ)*(NRBwt/BalabanJ)))	0.946
11	(((Ecc*ROG)+(FCSP3*BalabanJ))/((ROG/NRBwt)*(BalabanJ-MPC)))	0.946
12	(((Ecc*NRBwt)*(NRBwt-ROG))/((FCSP3-ROG)*sqrt(BalabanJ)))	0.945
13	(((FCSP3-NRBwt)/(FCSP3+BalabanJ))/((Ecc/NRBwt)+(VSA5/Chi2v)))	0.945
14	(((Ecc*MPC)*(ROG*NRBwt))/((FCSP3-ROG)*sqrt(BalabanJ)))	0.945
15	(((Ecc*NRBwt)/(FCSP3-ROG))/(sqrt(BalabanJ)/ROG))	0.945
16	(((ROG*NRBwt)/(FCSP3-ROG))*(exp(Ecc)/sqrt(BalabanJ)))	0.945
17	(((MPC*BalabanJ)/(BalabanJ-Ecc))/((FCSP3-ROG)/(ROG*NRBwt)))	0.945
18	(((Ecc+NRBwt)*(ROG*NRBwt))/((FCSP3-ROG)*sqrt(BalabanJ)))	0.945
19	(((FCSP3/ROG)+(Ecc/BalabanJ))*((NRBwt/MPC)+(NRBwt/BalabanJ)))	0.945
20	(((NRBwt*BalabanJ)/(BalabanJ-Ecc))/((FCSP3-ROG)/(NRBwt-ROG)))	0.945

Table S III. SISSO-identified top-20 descriptors used to quantify the relative feature importances in the present work.

Details of Kernel Ridge Regression (KRR): KRR is a similarity based regression algorithm that inherently takes the nonlinearity of the system into account. The 'similarity' between any two samples (in our case polymer compositions) is defined in the feature space using some standard mathematical measure of distance, such as a Euclidean distance. For any two polymers i and j having feature vectors f_i and f_j respectively (where f_i is an m dimensional vector with components $f_i^1, f_i^2, f_i^3 ... f_i^m$), the Euclidean distance between them

will be defined as:

$$d((f_i, f_j)) = \|f_i - f_j\|_2 = \sqrt{(f_i^1 - f_j^1)^2 + (f_i^2 - f_j^2)^2 + ... + (f_i^m - f_j^m)^2}. \quad (1)$$

The smaller (larger) is this distance in the m-dimensional feature space, the more similar (dissimilar) the two polymers are. Subsequently, the property of interest for a new polymer j, denoted by $P(j)$, can be written as follows:

$$P(j) = \sum_{i=1}^{n} \alpha_i \mathcal{K}(f_i, f_j). \quad (2)$$

Where the summation is performed over the entire training set size n, and $\mathcal{K}(f_i, f_j)$ is the *kernel function* that is defined in terms of $d(f_i, f_j)$. In the present case we used a Gaussian kernel, which is defined as:

$$\mathcal{K}_{\mathcal{G}}(f_i, f_j) = exp\left(\frac{-\|f_i - f_j\|_2^2}{2\sigma^2}\right). \quad (3)$$

Here σ is a kernel hyper-parameter known as the Gaussian width. The kernel coefficients α_i are determined as a part of the training procedure which essentially involves an iterative minimization of prediction errors leading to the optimal parameter choices, while following best practices of statistical learning such as cross-validation and regularization. Mathematically, the training process involves a minimization of the following expression:

$$\underset{\alpha_1,...,\alpha_n}{\arg\min} \sum_{i=1}^{n} (P_{pred}(i) - P_{actual}(i))^2 + \lambda \sum_{i=1}^{n} \|\alpha_i\|_2^2. \quad (4)$$

Where $P_{pred}(i)$ is the KRR model predicted property value of polymer i and $P_{actual}(i)$ is its actual property value; $(P_{pred}(i) - P_{actual}(i))$ is thus a measure of the prediction error. The second term in the expression involves the regularization parameter λ, which is an important step that prevents overfitting. The solution to the above minimization problem is analytically given by the following equation:

$$\alpha = (\mathbf{K}_{train} + \lambda \mathbf{I})^{-1} \mathbf{P}_{\mathbf{actual}} \quad (5)$$

Where α is the vector of all α_i values, $\mathbf{K}_{train}$ represents the kernel matrix for the entire training set, and $\mathbf{P}_{\mathbf{actual}}$ represents the vector of actual property values for all points in the training set. Thus, the two parameters that need to be optimized while the KRR model

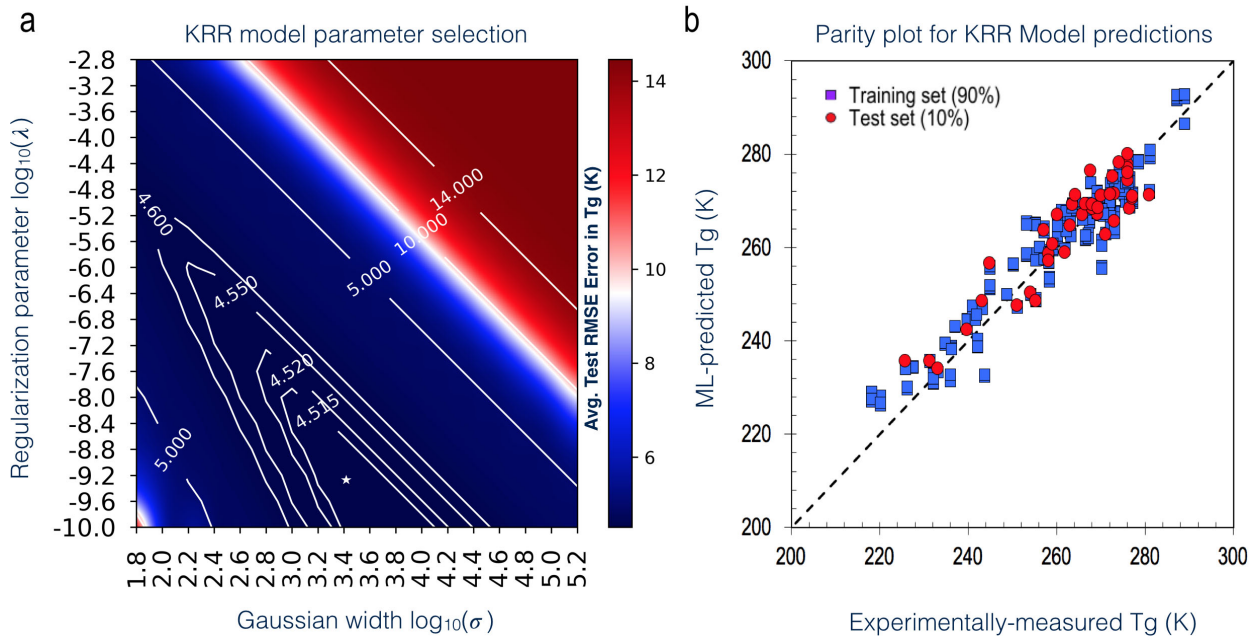

FIG. S1. KRR model's prediction performance on a a randomly-selected 10% left-out dataset as a function of different sets of model hyper-parameters, namely, the regularization parameter λ and the Gaussian width parameter σ. The optimal hyper-parameter set is identified by a '$\star$'. (b) A parity plot comparing KRR models training and test set prediction performances with the corresponding experimentally measured T_g values. Three different randomly-selected training and test splits are plotted to show that a similar performance is achieved with respect to different training/test splits.

building exercise are the regularization parameter λ and the Gaussian width parameter σ. The optimal choices identified for these parameters are indicates by a '$\star$' in Fig. S1(a). The final ML model built of the selected optimal hyper-parameter choices is then shown to make accurate T_g predictions on unseen data (*cf.*, Fig. S1(b)). We confirm that the test set prediction accuracy achieved with the KRR ML model is comparable to that of the RF-based ML model described in the main text of the manuscript.

---

* gpilania@lanl.gov

[1] K. Fritzsche, R. W. Lenz, and R. C. Fuller, Die Makromolekulare Chemie: Macromolecular Chemistry and Physics **191**, 1957 (1990).

[2] G. A. Abraham, A. Gallardo, J. San Roman, E. R. Olivera, R. Jodra, B. García, B. Miñambres,

J. L. García, and J. M. Luengo, Biomacromolecules **2**, 562 (2001), ISSN 1525-7797.

[3] S. Mizuno, S. Katsumata, A. Hiroe, and T. Tsuge, Polymer Degradation and Stability **109**, 379 (2014), ISSN 0141-3910.

[4] R. Luo, J. Chen, L. Zhang, and G. Chen, Biochemical Engineering Journal **32**, 218 (2006), ISSN 1369-703X.

[5] C.-K. Kang, S. Kusaka, and Y. Doi, Biotechnol Lett **17**, 583 (1995), ISSN 1573-6776.

[6] J. Myung, J. C. A. Flanagan, R. M. Waymouth, and C. S. Criddle, Process Biochemistry **51**, 561 (2016), ISSN 1359-5113.

[7] X. Han, Y. Satoh, T. Satoh, K. Matsumoto, T. Kakuchi, S. Taguchi, T. Dairi, M. Munekata, and K. Tajima, Appl Microbiol Biotechnol **92**, 509 (2011), ISSN 1432-0614.

[8] N. Tanadchangsaeng, A. Kitagawa, T. Yamamoto, H. Abe, and T. Tsuge, Biomacromolecules **10**, 2866 (2009), ISSN 1525-7797.

[9] H.-h. Wang, X.-r. Zhou, Q. Liu, and G.-Q. Chen, Appl Microbiol Biotechnol **89**, 1497 (2011), ISSN 1432-0614.

[10] Y.-H. Na, Y. He, N. Asakawa, N. Yoshie, and Y. Inoue, Macromolecules **35**, 727 (2002), ISSN 0024-9297.

[11] E. Shimamura, M. Scandola, and Y. Doi, Macromolecules **27**, 4429 (1994), ISSN 0024-9297.

[12] S. Chanprateep and S. Kulpreecha, Journal of Bioscience and Bioengineering **101**, 51 (2006), ISSN 1389-1723.

[13] B. Laycock, M. V. Arcos-Hernandez, A. Langford, J. Buchanan, P. J. Halley, A. Werker, P. A. Lant, and S. Pratt, Journal of Applied Polymer Science **131** (2014), ISSN 1097-4628.

[14] R. D. Ashby, D. K. Y. Solaiman, A. Nuñez, G. D. Strahan, and D. B. Johnston, Bioresource Technology **253**, 333 (2018), ISSN 0960-8524.

[15] Y. Dai, Z. Yuan, K. Jack, and J. Keller, Journal of Biotechnology **129**, 489 (2007), ISSN 0168-1656.

[16] A. Ferre-Guell and J. Winterburn, Biomacromolecules **19**, 996 (2018), ISSN 1525-7797.

[17] J. Han, L.-P. Wu, J. Hou, D. Zhao, and H. Xiang, Biomacromolecules **16**, 578 (2015), ISSN 1525-7797.

[18] T. Huu Phong, D. Van Thuoc, and K. Sudesh, International Journal of Biological Macromolecules **84**, 361 (2016), ISSN 0141-8130.

[19] S. Povolo, M. G. Romanelli, M. Basaglia, V. I. Ilieva, A. Corti, A. Morelli, E. Chiellini, and

S. Casella, New Biotechnology **30**, 629 (2013), ISSN 1871-6784.

[20] K.-H. Huong, M. J. Azuraini, N. A. Aziz, and A.-A. A. Amirul, Journal of Bioscience and Bioengineering **124**, 76 (2017), ISSN 1389-1723.

[21] I. M. Syafiq, K.-H. Huong, K. Shantini, S. Vigneswari, N. A. Aziz, A.-A. A. Amirul, and K. Bhubalan, Enzyme and Microbial Technology **98**, 1 (2017), ISSN 0141-0229.

[22] N. M. Iqbal and A. A. Amirul, Journal of Chemical Technology & Biotechnology **89**, 407 (2014), ISSN 1097-4660.

[23] N. Zhila and E. Shishatskaya, International Journal of Biological Macromolecules **111**, 1019 (2018), ISSN 0141-8130.

[24] J. Zhang, K. Kasuya, T. Hikima, M. Takata, A. Takemura, and T. Iwata, Polymer Degradation and Stability **96**, 2130 (2011), ISSN 0141-3910.

[25] D. Hu, A.-L. Chung, L.-P. Wu, X. Zhang, Q. Wu, J.-C. Chen, and G.-Q. Chen, Biomacromolecules **12**, 3166 (2011), ISSN 1525-7797.

[26] X.-W. Shen, Y. Yang, J. Jian, Q. Wu, and G.-Q. Chen, Bioresource Technology **100**, 4296 (2009), ISSN 0960-8524.

[27] L. Tripathi, L.-P. Wu, M. Dechuan, J. Chen, Q. Wu, and G.-Q. Chen, Bioresource Technology **142**, 225 (2013), ISSN 0960-8524.

[28] D.-C. Meng, Z.-Y. Shi, L.-P. Wu, Q. Zhou, Q. Wu, J.-C. Chen, and G.-Q. Chen, Metabolic Engineering **14**, 317 (2012), ISSN 1096-7176.

[29] L. Tripathi, L.-P. Wu, D. Meng, J. Chen, and G.-Q. Chen, Biomacromolecules **14**, 862 (2013), ISSN 1525-7797.

[30] L. H. Hall and L. B. Kier, in *Reviews in Computational Chemistry* (John Wiley & Sons, Ltd, 2007), pp. 367–422, ISBN 978-0-470-12579-3.

[31] A. T. Balaban, Chemical Physics Letters **89**, 399 (1982), ISSN 0009-2614.

[32] G. Landrum, *RDKit: Open-Source Cheminformatics http://www. Rdkit. Org* (Accessed, 2016).

[33] P. Labute, Journal of Molecular Graphics and Modelling **18**, 464 (2000), ISSN 1093-3263.

[34] S. A. Wildman and G. M. Crippen, J. Chem. Inf. Comput. Sci. **39**, 868 (1999), ISSN 0095-2338.

[35] *UCSF DOCK 3.7*, `http://dock.compbio.ucsf.edu/DOCK3.7/` (2019).

[36] *Chemical Computing Group (CCG) Computer-Aided Molecular Design*, `https://www.chemcomp.com/` (2019).

[37] *RDKit: Open-source cheminformatics*, `https://www.rdkit.org/` (2019).

[38] R. Todeschini and V. Consonni, in *Handbook of Chemoinformatics* (John Wiley & Sons, Ltd, 2008), pp. 1004–1033, ISBN 978-3-527-61827-9.

[39] W. H. B. Sauer and M. K. Schwarz, J. Chem. Inf. Comput. Sci. **43**, 987 (2003), ISSN 0095-2338.

[40] G. A. Arteca, in *Reviews in Computational Chemistry* (John Wiley & Sons, Ltd, 2007), pp. 191–253, ISBN 978-0-470-12586-1.