Development of Complete Enumeration Program for Polymer Structural Isomers: A Case of Nylon-*n* (*n* = 3 - 10)

Gikyo Usuki[1], Takuro Tsutsumi[2], Kohei Hasebe[2], Masato Kobayashi[2,3], Keitaro Matsuoka[2], Tetsuya Taketsugu[2,3], and Kazuki Sada[2]

[1] Graduate School of Chemical Sciences and Engineering, Hokkaido University, Sapporo 060-0810, Japan 94378

[2] Department of Chemistry, Faculty of Science, Hokkaido University, Sapporo 060-0810, Japan 94378

[3] Institute for Chemical Reaction Design and Discovery (WPI-ICReDD), Hokkaido University, Sapporo 001-0021, Japan 94378

*Corresponding author: Kazuki Sada

Email: sadatcm@sci.hokudai.ac.jp

Keywords: Polymer structural isomer, Virtual polymer, Complete enumeration, Nylon

Abstract

The enumeration of structural isomers of organic molecules using programs capable of enumerating all structural isomers for a given molecular formula has been expanding the chemical library. Here, we report the development of the program (PolyENU) that completely enumerates polymer structural isomers. PolyENU program enumerates all isomers of the repeating unit defined by a molecular formula and a functional group constituting the polymer structure. From the repeating unit of nylon 6 ($C_6H_{11}NO$ and an amide group), 387 polymer isomers were generated by PolyENU. The enumerated polymer library includes common organic polymers such as nylon 6, polyleucine, poly(2-isopropyl-oxazoline), and poly(*N*-propylacrylamide), as well as many unexplored polymers. The enumeration of nylon *n* (*n* = 3 - 10) yielded 36,505 polymers within 44 minutes. These results clearly indicate that PolyENU expands polymer chemical space and contributes to diversity in polymer science.

1. Introduction

Isomers in organic chemistry refer to compounds with the same molecular formula but different molecular structures. As the number and variety of constituent atoms in a molecule increase, the number of isomers exponentially rises, ensuring the diversity of organic molecules. The exploration of structural isomers began classically with alkanes and has evolved to modern informatics, employing combinatorial theory to develop programs capable of enumerating all structural isomers for a given molecular formula.[1–5] Presently, approximately 166 billion isomers of organic molecules up to 17 atoms have been elucidated.[6]

On another front, polymer chemistry traces its roots back to Staudinger's conceptualization of polymers[7] and Carothers' pioneering work in synthetic rubbers and fibers.[8,9] Since then, polymer scientists have been developing new polymers and seeking new functions of the polymers through huge experimental studies. Recently, the improvement of the functionality of the polymers has been accelerated through machine learning.[10–12] In particular, the exploration of polymer chemical space has been expanding with the construction of huge experimental libraries from the existing polymers[13] or virtual libraries from commercially available monomers synthesized by the standard protocols.[14,15] These libraries are leveraged to facilitate the exploration of polymers and their functions. Here, we report on the development of the enumeration program of all structural isomers of polymers. Disregarding the availability of monomers and the polymerization protocols, we offered a comprehensive set of polymers defined by the molecular formula and the functional group. In this report, we applied it to the complete enumeration of the repeating units of a series of nylons.

2. Enumeration of Polymer Structural Isomers

Figure 1 shows the scheme for the enumeration of polymer structural isomers, which is named PolyENU. PolyENU is a SMILES-based Python program that allows the generation of all polymer structural isomers in the polymer chemical space from a given repeating unit defined by the molecular formula and the functional group. PolyENU first requires these two structural parameters of the polymer for the exploration of polymer chemical space. **Figure 1** shows an example of the polymer enumeration defined by $C_6H_{11}NO$ for the molecular formula and an amide group for the functional group. PolyENU then generates an exhaustive set of substituent groups that can be introduced into the functional group by determining the atom sets usable as substituents, which are the remaining atoms from the given molecular formula minus the partial molecular formula corresponding to the functional group. Note that two polymerization points must be added to represent linear polymers.

Next, the generation of substituent candidates is performed using *enu*[5], a program for the complete enumeration of structural isomers for small molecules. The *enu* program has so far succeeded in enumerating all structural isomers from CH_4 to $C_{25}H_{52}$[5]. In PolyENU, substituent candidates are systematically generated based on the set of usable atoms for the substituent, which also includes the polymerization and substitution points. The examples of substituent candidates are summarized in **Table 1**. After that, polymer structural isomers are listed by exploring all pairs of substituents that satisfy the given molecular formula with two polymerization points. At present, only chain polymers with two polymerization points can be generated.

The enumerated SMILES data for structural isomers may be duplicated due to symmetrically located substitution points and ambiguity in the definition of repeating unit structures. For small molecules, canonical SMILES, a unique representation of SMILES, can be

used to remove duplicate structures. Recently, BigSMILES[16], an extension of SMILES to polymers, and its canonical representation[17] have been proposed, but the source code is not publicly available. In order to remove duplicate structures, PolyENU utilizes a canonical representation of cyclic small molecules converted from repeating unit structures by directly closing the two polymerization points. Since the smallest unit of cyclic molecules is a three-membered ring, polymers with one or two atoms in the main chain are trimerized or dimerized, respectively, before the ring closure. By following the above scheme, PolyENU completely enumerates polymer structural isomers defined by any molecular formula and a functional group.

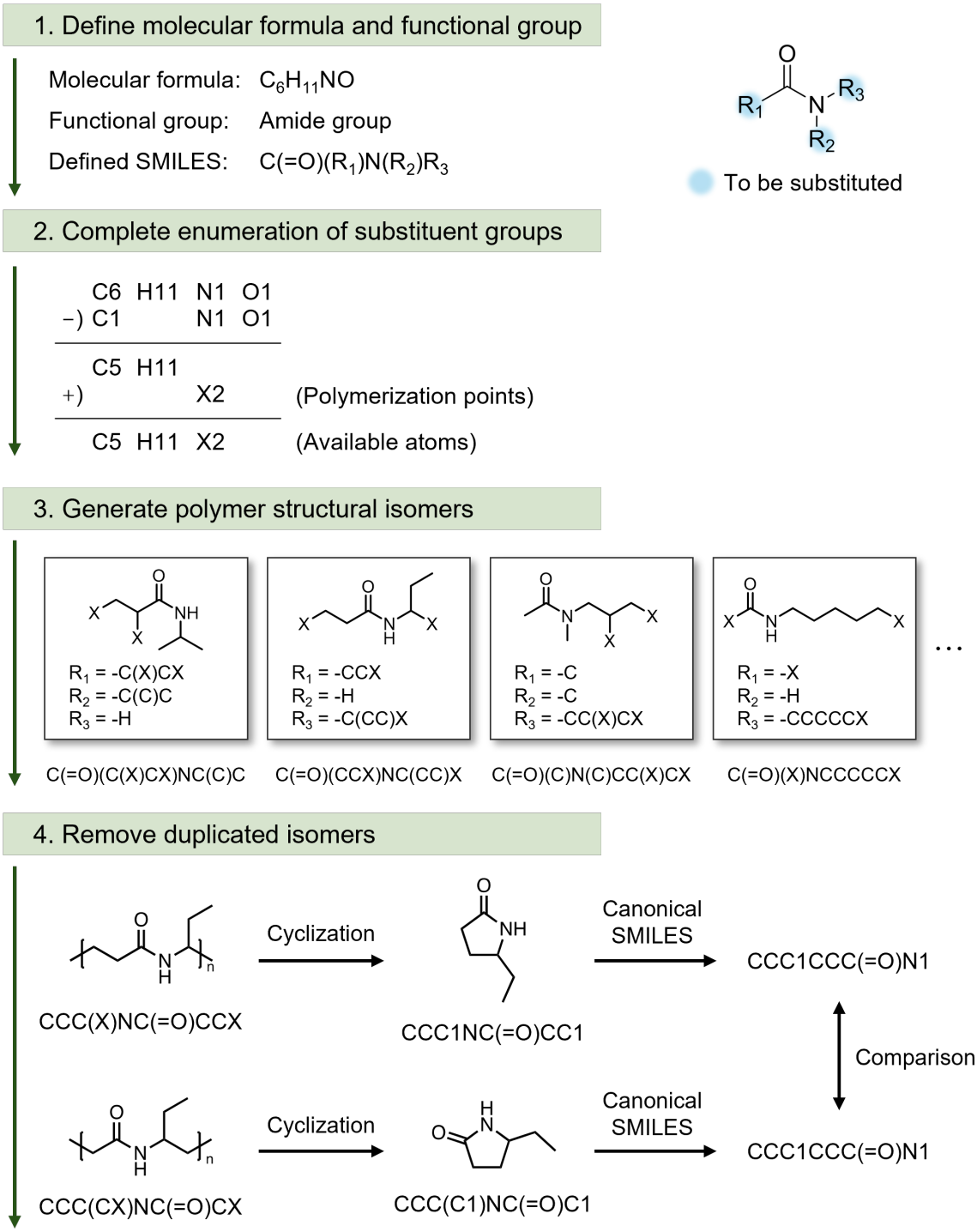

Figure 1. PolyENU scheme for enumeration of polymer structural isomers.

Table 1. Examples of substituent candidates enumerated by the *enu* program[5].

# of C	Substituent candidates
0	-H, -X
1	-C, -CX, -C(X)X
2	-CC, -C(X)C, -CC(X)X, ...
3	-C(C)C, -C(C)CX, -C(X)CCX, ...
4	-C(C)(X)CC, -CC(X)CX, ...
5	-C(CC)CC, -C(CC)(CX)CX, ...

3. Results and Discussion

3.1. Enumeration of polymer structural isomers of nylon 6

To demonstrate the PolyENU program, we addressed the enumeration of structural isomers of nylon 6 (poly(ε-caprolactam)), a typical synthetic fiber, using $C_6H_{11}NO$ as the molecular formula and an amide group as the functional group as the input parameters of the repeating unit, yielding 387 polymer isomers. Among them, representative examples are shown in **Figure 2a**. They have diversity beyond the classifications in polymer science, such as nylon 6 (**1**), polyleucine (**2**), poly(*N*-pentyl)isocyanate (**3**), poly(2-isopropyl-oxazoline) (**4**), poly(*N*-propylacrylamide) (**5**), poly(*N*-ethylmethacrylamide) (**6**), and poly(*N*-vinylbutyramide) (**7**). Thermo-responsive polymers in water, such as poly(*N*-isopropyl acrylamide) (PNIPAM)[18] and poly(2-isopropyl-2-oxazoline)[19] are included. Therefore, many kinds of polymers with various physical properties are depicted in the complete enumeration, and many unexplored polymers are listed. This indicates that the enumeration expands polymer chemical space.

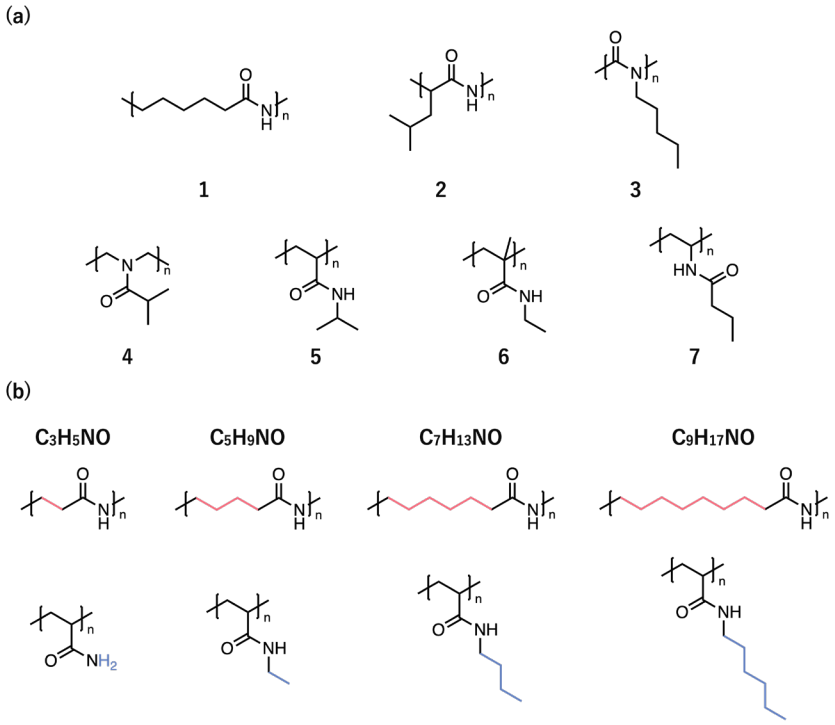

Figure 2. Representative examples of the polymer structural isomers in the $C_6H_{11}NO$ polymer group. (a) Commodity plastics and an unexplored polymer. (b) polymer analogs.

3.2. Systematic enumeration of polymer structural isomers of nylon *n* (*n* = 3-10)

For further demonstration, we systematically enumerated polymer structural isomers for nylon *n* (*n* = 3 - 10). **Figure 2b** shows examples of polymer analogs related to the main chain (top) and the side chain (bottom) structures, respectively. PolyENU is a complete enumeration program that restricts a molecular formula and a functional group composing repeating unit structures, thus making it directly impossible to obtain polymer analogs. However, it was shown that PolyENU can enumerate not only polymer structural isomers but also polymer analogs by systematically changing the number of atoms, especially the carbon atom, in the molecular formula.

Table 2 summarizes the total number of enumerated structural isomers of nylon *n* (*n* = 3 - 10) and the computational time, together with the number of polymer structural isomers as a function of the number of atoms in the main chain. The number of polymer isomers increases as the total number of atoms in the molecular formula increases. Moreover, we also estimated the numbers of the small molecules generated from the repeating unit terminated by two hydrogen atoms. The increase of the total numbers observed in the nylon *n* is considerably lower than that in small molecule analog: 23,416 isomers for the polymer with $C_{10}H_{19}NO$ and 330,823 for the small molecules with $C_{10}H_{21}NO$. This is attributed to many duplications in the enumeration of the polymers due to symmetrically located substitution points and ambiguity in the definition of repeating unit. In addition, PolyENU achieves complete enumeration in practical time using the *enu* program, which is capable of very fast enumeration through graph theory. In fact, it confirmed that 36,505 polymers in the polymer chemical space of nylon *n* (*n* = 3 - 10) could be generated within approximately 44 minutes. Most industrially used commodity plastics have roughly 40 or fewer atoms in their repeating unit structure, which suggests that the PolyENU program can be applied to commodity polymers with various functions for further exploration of new polymers.

Table 2. Summary of complete enumerations of polymer structural isomers for $C_nH_{2n-1}NO$ homopolymer groups (n = 3 - 10).

# of atoms in main chain	C_3H_5NO (n = 3)	C_4H_7NO (n = 4)	C_5H_9NO (n = 5)	$C_6H_{11}NO$ (n = 6)	$C_7H_{13}NO$ (n = 7)	$C_8H_{15}NO$ (n = 8)	$C_9H_{17}NO$ (n = 9)	$C_{10}H_{19}NO$ (n = 10)	Total
1	8	21	55	146	388	1,036	2,778	7,479	11,911
2	5	15	42	116	319	876	2,397	6,561	10,331
3	3	8	22	61	168	462	1,271	3,494	5,489
4	1	4	13	39	115	332	948	2,686	4,138
5	-	1	5	18	58	179	538	1,588	2,387
6	-	-	1	6	25	89	295	936	1,352
7	-	-	-	1	7	32	124	438	602
8	-	-	-	-	1	8	41	173	223
9	-	-	-	-	-	1	9	50	60
10	-	-	-	-	-	-	1	10	11
11	-	-	-	-	-	-	-	1	1
Total	17	49	138	387	1,081	3,015	8,402	23,416	36,505
[a]Time / s	1.4	2.6	6.3	18	50	168	542	1,860	2,648.3 (~ 44 m)
[b]Small molecules	84	299	1,015	3,345	10,777	34,156	106,849	330,823	487,348

[a]Computed by a workstation-type Intel Xeon Gold 6136 3.00 GHz processor with only one core. The memory capacity was 192 GB. [b]Total number of complete enumerations for small molecule generated from the repeating unit terminated by two hydrogen atoms ($C_nH_{2n+1}NO$).

4. Conclusion

We have developed PolyENU, a SMILES-based Python program for exhaustively enumerating polymer structural isomers. PolyENU performs the complete enumeration of polymers in the polymer chemical space based on a molecular formula and a single functional group of a repeating unit structure. As the first demonstration, many kinds of polymers are generated, such as polyamides, polypeptides, and polyacrylamides as the structural isomers of nylon 6. We addressed the complete enumeration of nylon *n* (*n* = 3 - 10) and obtained 36,505 different structural isomers within 44 minutes. In conclusion, the expanded polymer chemical space generated by PolyENU contributes to diversity in polymer science.

References

(1) Molchanova, M. S.; Shcherbukhin, V. V; Zefirov, N. S. *J. Chem. Inf. Comput. Sci.* **1996**, 36, 888.

(2) Melnikov, A. A.; Palyulin, V. A.; Zefirov, N. S. *J. Chem. Inf. Model.* **2007**, 47, 2077.

(3) Fink, T.; Bruggesser, H.; Reymond, J. L. *Angew. Chem. Int. Ed.* **2005**, 44, 1504.

(4) Blum, L. C.; Reymond, J. L. *J. Am. Chem. Soc.* **2009**, 131, 8732.

(5) Rieder, S. R.; Oliveira, M. P.; Riniker, S.; Hünenberger, P. H. *J. Cheminform.* **2023**, 15.

(6) Ruddigkeit, L.; Van Deursen, R.; Blum, L. C.; Reymond, J. L. *J. Chem. Inf. Model.* **2012**, 52, 2864.

(7) Staudinger, H. *Eur. J. Inorg. Chem.* **1920**, 53, 1073.

(8) Carothers, W. H.; Williams, I.; Collins, A. M.; Kirby, J. E. *J. Am. Chem. Soc.* **1931**, 53, 4203.

(9) Carothers, W. H. US2130523A. **1938**.

(10) Duchowicz, P. R.; Fioressi, S. E.; Bacelo, D. E.; Saavedra, L. M.; Toropova, A. P.; Toropov, A. A. *Chemometr. Intell. Lab. Syst.* **2015**, 140, 86.

(11) Cravero, F.; Díaz, M. F.; Ponzoni, I. *J. Chem. Phys.* **2022**, 156.

(12) Wu, J. Q.; Gong, X. Q.; Wang, Q.; Yan, F.; Li, J. J. *Chem. Eng. Sci.* **2023**, 267.

(13) Otsuka, S.; Kuwajima, I.; Hosoya, J.; Xu, Y.; Yamazaki, M. *IEEE*. **2011**, 22.

(14) Ohno, M.; Hayashi, Y.; Zhang, Q.; Kaneko, Y.; Yoshida, R. *J. Chem. Inf. Model.* **2023**, 63, 5539.

(15) Hayashi, Y.; Shiomi, J.; Morikawa, J.; Yoshida, R. *NPJ Comput. Mater.* **2022**, 8.

(16) Lin, T. S.; Coley, C. W.; Mochigase, H.; Beech, H. K.; Wang, W.; Wang, Z.; Woods, E.; Craig, S. L.; Johnson, J. A.; Kalow, J. A.; Jensen, K. F.; Olsen, B. D. *ACS Cent. Sci.* **2019**, 5, 1523.

(17) Lin, T. S.; Rebello, N. J.; Lee, G. H.; Morris, M. A.; Olsen, B. D. *ACS Polym. Au* **2022**, 2, 486.

(18) Halperin, A.; Kröger, M.; Winnik, F. M. *Angew. Chem. Int. Ed.* **2015**, 54, 15342.

(19) Hoogenboom, R.; Schlaad, H. *Polym. Chem.* **2016**, 8, 24.

Acknowledgments

This work was partly supported by JST, ACT-X Grant Number JPMJAX23DE, Japan for TT, by JSPS KAKENHI Grant Number JP21H01980 for KS, and by Promotion for Young Research Talent and Network from Northern Advancement Center for Science & Technology (NOASTEC) of Hokkaido Japan for KM.

Author Contributions

G. Usuki, **T. Tsutsumi**: Conceptualization, Data curation, Methodology, Investigation, Software, Visualization, Writing – original draft, Writing – review & editing.

K. Hasebe, **M. Kobayashi**, **K. Matsuoka**: Conceptualization, Methodology, Software, Writing – original draft, Writing – review & editing.

T. Taketsugu, **K. Sada**: Conceptualization, Methodology, Visualization, Writing – original draft, Writing – review & editing.

Competing Interests

The authors declare no competing interests.

ORCID

Takuro Tsutsumi: 0000-0001-5086-1122

Masato Kobayashi: 0000-0002-4001-3581

Keitaro Matsuoka: 0000-0002-6213-1220

Tetsuya Taketsugu: 0000-0002-1337-6694