

Supplementary Information for: “SAVI, in silico generation of billions of easily synthesizable compounds through expert-system type rules”

Hitesh Patel¹, Wolf-Dietrich Ihlenfeldt², Philip N. Judson³, Yurii Moroz⁴, Yuri Pevzner^{1,5}, Megan L. Peach⁶, Victorien Delannée¹, Nadya Tarasova⁷, Marc C. Nicklaus¹

¹*Computer-Aided Drug Design Group, Chemical Biology Laboratory, Center for Cancer Research, National Cancer Institute, National Institutes of Health, Frederick, MD 21702, USA.*

²*Xemistry GmbH, Schwalbenweg 5, D-61479 Glashütten, Germany.*

³*Heather Lea, Bland Hill, Norwood, Harrogate HG3 1TE, England.*

⁴*Enamine Ltd, 78 Chervonotkatska Street, Suite 1, Kyiv, 02094, Ukraine and Chemspace LLC, 85 Chervonotkatska Street, Suite 1, Kyiv, 02094, Ukraine*

⁵*AbbVie, Inc., North Chicago, IL 60064, USA*

⁶*Basic Science Program, Frederick National Laboratory for Cancer Research, Frederick, MD 21702*

⁷*Synthetic Biologics and Drug Discovery Group, Laboratory of Cancer Immunometabolism, Center for Cancer Research, National Cancer Institute, National Institutes of Health, Frederick, MD 21702, USA.*

Table of Contents

1. Products Properties List.....	3
2. Reactions Properties List	4
3. Fields of SD file of a SAVI product.....	6
4. Additional plots of property distribution for “Plus” set	9
5. Plots of property distribution for “Neg0” set	11
6. Plots of property distribution for “Neg10” set	14
7. Plots of property distribution for “Neg20” set	17
8. Plots of property distribution for “Neg30” set	20

1. Products Properties List

SAVI_ID - Product name, SAVI ID

HASHCODE - Product Hashcode

SMILES - Product SMILES

FORMULA - Molecular formula

WEIGHT - Molecular weight

NROTBONDS - Number of rotatable bonds

NHDONORS - Number of H-bond donors

NHACCEPTORS - Number of H-bond acceptors

TPSA - Total polar surface area

FSP3 - Fraction of sp³-hybridized carbons

COMPLEXITY - Complexity (see: W. D. Ihlenfeldt, Ph.D. Thesis, TU Munich, Germany, 1991)

CHEMICAL_BEAUTY - QED score of the product

BENZENOID_INDEX - Benzenoid index

CHARGED_GROUP_COUNTS - Number of charged groups

HYDROGEN_BOND_CENTER_COUNT - Number of hydrogen bond centers

HEAVY_ATOM_COUNT - Number of heavy atoms

STEREO_COUNT - Number of stereo centers

RINGSYSTEM_COUNT - No. of ring systems

NESSSR - Number of extended smallest set of smallest rings (ESSSR)

NSSSR - Number of smallest set of smallest rings (SSSR)

ARORING_COUNT - No. of aromatic rings

ALIRING_COUNT - No. of aliphatic rings

STDINCHI - Standard InChI

STDINCHIKEY - Standard InChI-Key

STEREO_TAUTO_HASH - tautomer-insensitive hashcode

RULE_OF_5_VIOLATIONS - Number of Lipinski "rule of 5" violations

RULE_OF_3_VIOLATIONS - Number of "rule of 3 violations"

SAVI_PAINS_FILTER_MATCH_COUNT - PAINS Filter Match

SAVI_PAINS_FILTER_MATCH_NAME - Name of the PAINS filter match

GENOTOXIC_ALERTS - Genotoxic alerts (see: CE&N Sep 27, 2010, p16ff)

XLOGP2 - Xlogp2

XLOGP - Xlogp

SAVI_BRUNS_WATSON_DEMERIT_SCORE - Bruns and Watson Demerit score (see Bruns, R.F., Watson, I.A. Rules for identifying potentially reactive or promiscuous compounds. J. Med. Chem. 2012, 55,9763-9772)

SAVI_BRUNS_WATSON_DEMERIT_COMPONENTS - Breakdown of Bruns and Watson Demerits

2. Reactions Properties List

SAVI_PROPOSED_TRANSFORM - Name of the reaction

SAVI_PROPOSED_TRANSFORM_ID - ID of the reaction (LHASA ID of the transform that describes this reaction)

SAVI_REACTION_HASHCODE - Hashcode of the reaction

SAVI_REACTION_SMILES - Atom-mapped SMILES of the reaction

SAVI_LHASA_SCORE - "Quality" score of this reaction according to LHASA transform scoring scheme

SAVI_LHASA_SCORING_DELTAPOINTS - Score penalties or rewards given to the reaction

SAVI_LHASA_SCORING_CLASS - Scoring class of the reaction according to highest delta points

SAVI_REACTION_CONDITIONS - Predicted reaction conditions according to LHASA transform rules

SAVI_REACTION_ACTUAL_CONDITIONS - Reaction conditions according to LHASA transform rules

SAVI_REACTION_WARNINGS - Reaction warnings according to LHASA transform rules

SAVI_SSSR_COUNT_CHANGE - SSSR count in product minus combined SSSR count in building blocks

SAVI_ALIRING_COUNT_CHANGE - No. of aliphatic rings changes in reaction

SAVI_ARORING_COUNT_CHANGE - No. of aromatic rings changes in reaction

SAVI_RINCHI - RInChI of the reaction

SAVI_LONG_RINCHIKEY - Long-RInChIKey of the reaction

SAVI_SHORT_RINCHIKEY - Short-RInChIKey of the reaction

SAVI_WEB_RINCHIKEY - Web-RInChIKey of the reaction

SAVI_BUILDING_BLOCK_A_HASHISY - Hashcode of building block A

SAVI_BUILDING_BLOCK_A_SMILES - SMILES of building block A

SAVI_BUILDING_BLOCK_A_BBSOURCE - Source with version of building block A (ENAMINE2019Q4 for this whole SAVI)

SAVI_BUILDING_BLOCK_A_STRID - Enamine catalog ID of building block A

SAVI_BUILDING_BLOCK_A_INCHI - InChI of building block A

SAVI_BUILDING_BLOCK_A_INCHIKEY - InChIKey of building block A

SAVI_BUILDING_BLOCK_A_PROTECTION_NEEDED - Indicates whether protection of reagent A is required in this reaction

SAVI_BUILDING_BLOCK_A_PROTECTED - Indicates whether building block A used in this reaction is already a protected version of the required reagent

SAVI_BUILDING_BLOCK_A_ORDER_LINK - URL of the Enamine catalog web page for building block A

SAVI_BUILDING_BLOCK_B_HASHISY - Hashcode of building block B

SAVI_BUILDING_BLOCK_B_SMILES - SMILES of building block B

SAVI_BUILDING_BLOCK_B_BBSOURCE - Source with version of building block B (ENAMINE2019Q4 for this whole SAVI)

SAVI_BUILDING_BLOCK_B_STRID - Enamine catalog ID of building block B

SAVI_BUILDING_BLOCK_B_INCHI - InChI of building block B

SAVI_BUILDING_BLOCK_B_INCHIKEY - InChIKey of building block B

SAVI_BUILDING_BLOCK_B_PROTECTION_NEEDED - Indicates whether protection of reagent B is required in this reaction

SAVI_BUILDING_BLOCK_B_PROTECTED - Indicates whether building block B used in this reaction is already a protected version of the required reagent

SAVI_BUILDING_BLOCK_B_ORDER_LINK - URL of the Enamine catalog web page for building block B

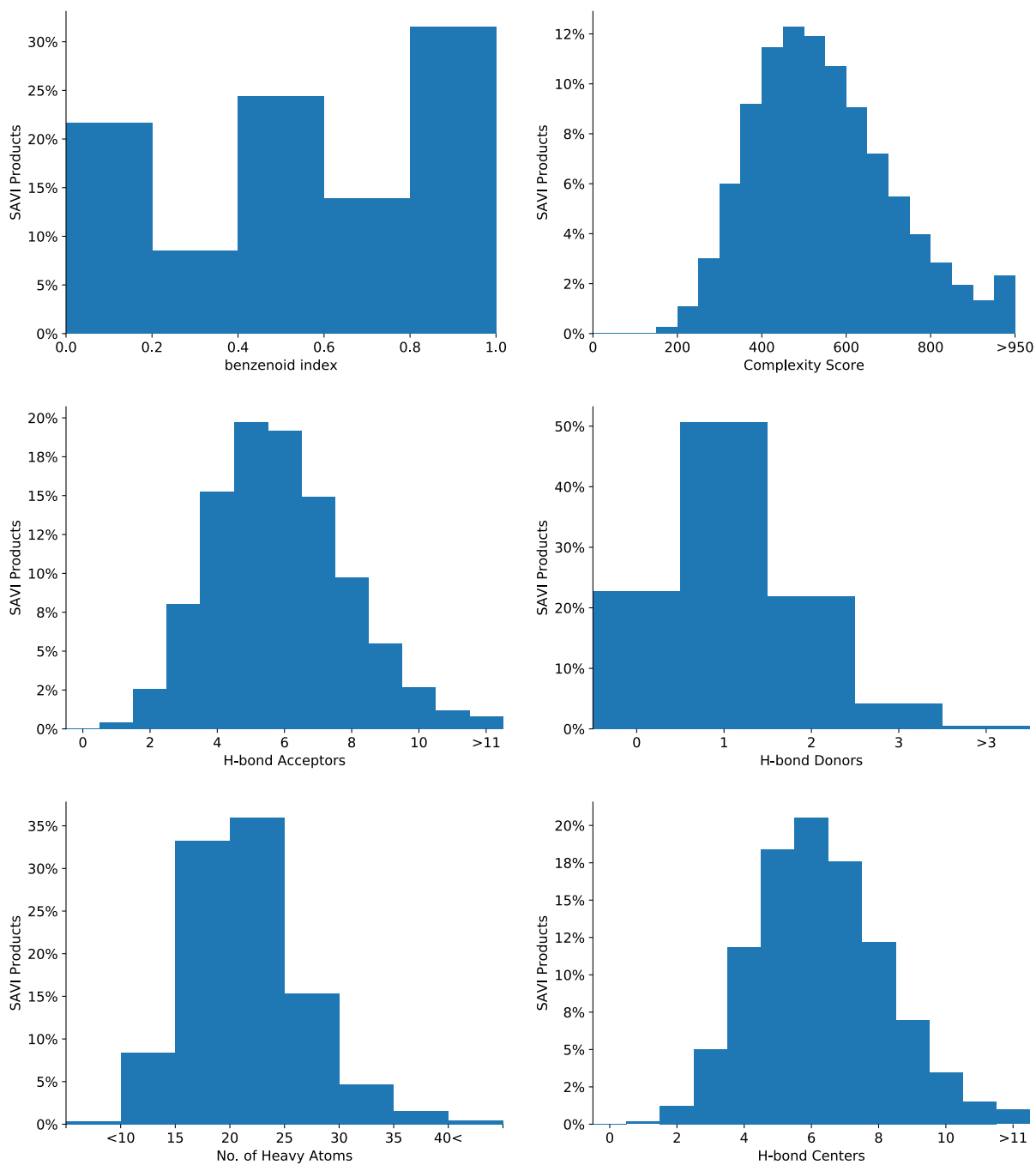
3. Fields of SD file of a SAVI product

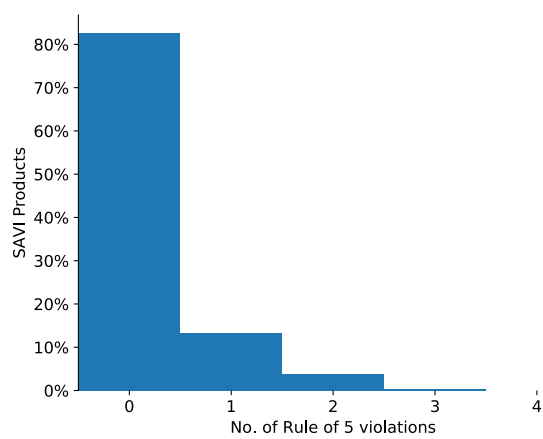
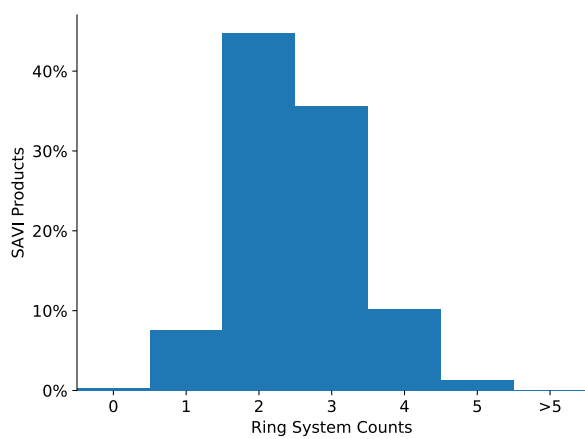
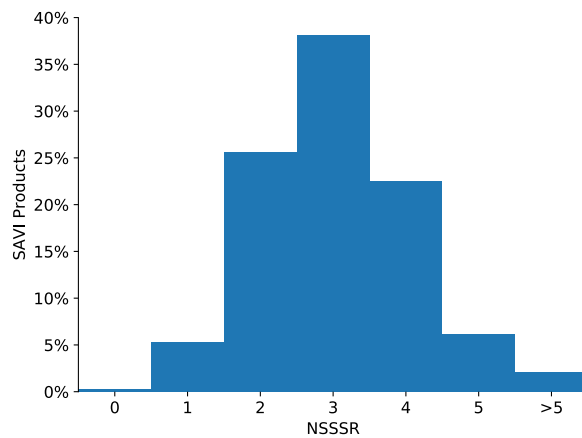
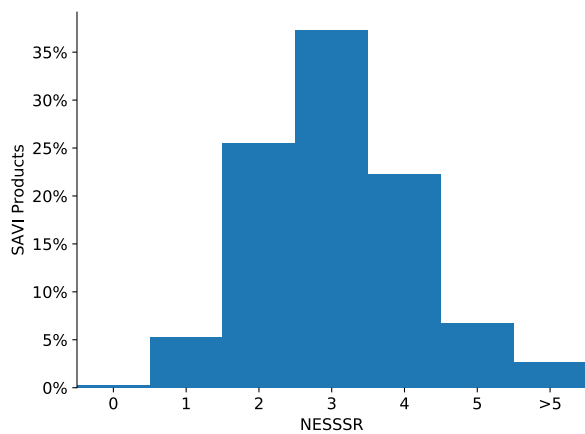
><E_SAVI_ID>
><E_SMILES>
><E_HASHISY>
><E_FORMULA>
><E_WEIGHT>
><E_NROTBONDS>
><E_NHDONORS>
><E_NHACCEPTORS>
><E_SAVI_PROTECTED>
><E_TPSA>
><E_FSP3>
><E_COMPLEXITY>
><E_CHEMICAL_BEAUTY>
><E_BENZENOID_INDEX>
><E_CHARGED_GROUP_COUNTS>
><E_HYDROGEN_BOND_CENTER_COUNT>
><E_HEAVY_ATOM_COUNT>
><E_STEREO_COUNT>
><E_RINGSYSTEM_COUNT>
><E_NESSSR>
><E_NSSSR>
><E_ALIRING_COUNT>
><E_ARORING_COUNT>
><E_SAVI_PAINS_FILTER_MATCH_COUNT>
><E_SAVI_PAINS_FILTER_MATCH_NAME>
><E_RULE_OF_5_VIOLATIONS>
><E_GENOTOXIC_ALERTS>
><E_XLOGP2>
><E_STEREO_TAUTO_HASH>
><E_SAVI_BRUNS_WATSON_DEMERIT_SCORE>

><E_SAVI_BRUNS_WATSON_DEMERIT_COMPONENTS>
><E_SAVI_PROPOSED_TRANSFORM>
><E_SAVI_PROPOSED_TRANSFORM_ID>
><E_SAVI_REACTION_HASHCODE>
><E_SAVI_REACTION_SMILES>
><E_SAVI_LHASA_SCORE>
><E_SAVI_LHASA_SCORING_DELTAPOINTS>
><E_SAVI_LHASA_SCORING_CLASS>
><E_SAVI_REACTION_CONDITIONS>
><E_SAVI_REACTION_ACTUAL_CONDITIONS>
><E_SAVI_REACTION_WARNINGS>
><E_SAVI_SSSR_COUNT_CHANGE>
><E_SAVI_ALIRING_COUNT_CHANGE>
><E_SAVI_ARORING_COUNT_CHANGE>
><E_SAVI_RINCHI>
><E_SAVI_LONG_RINCHIKEY>
><E_SAVI_SHORT_RINCHIKEY>
><E_SAVI_WEB_RINCHIKEY>
><E_SAVI_BUILDING_BLOCK_A_HASHISY>
><E_SAVI_BUILDING_BLOCK_A_SMILES>
><E_SAVI_BUILDING_BLOCK_A_BBSOURCE>
><E_SAVI_BUILDING_BLOCK_A_STRID>
><E_SAVI_BUILDING_BLOCK_A_INCHI>
><E_SAVI_BUILDING_BLOCK_A_INCHIKEY>
><E_SAVI_BUILDING_BLOCK_A_PROTECTION_NEEDED>
><E_SAVI_BUILDING_BLOCK_A_PROTECTED>
><E_SAVI_BUILDING_BLOCK_A_ORDER_LINK>
><E_SAVI_BUILDING_BLOCK_B_HASHISY>
><E_SAVI_BUILDING_BLOCK_B_SMILES>
><E_SAVI_BUILDING_BLOCK_B_BBSOURCE>
><E_SAVI_BUILDING_BLOCK_B_STRID>
><E_SAVI_BUILDING_BLOCK_B_INCHI>

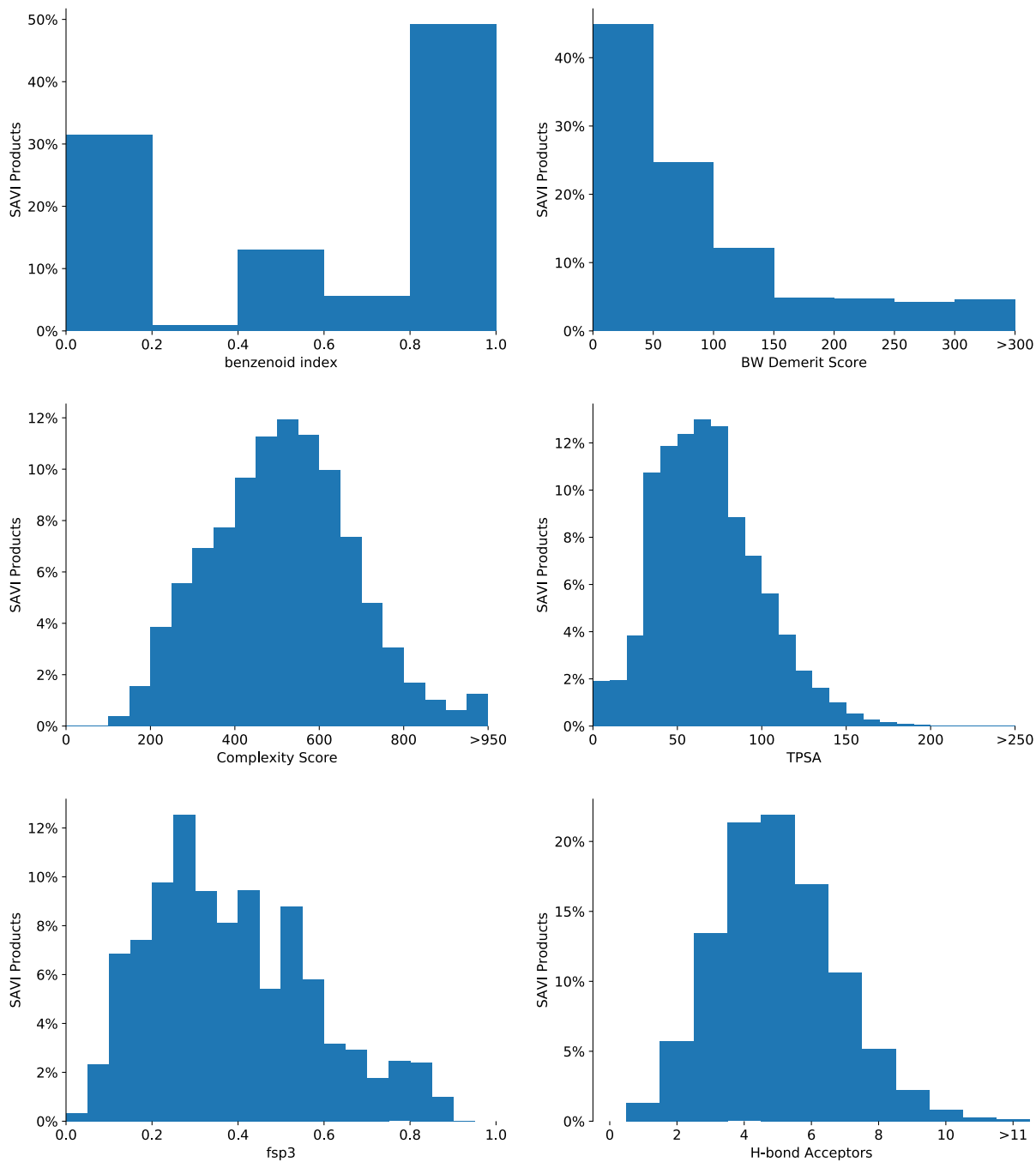
><E_SAVI_BUILDING_BLOCK_B_INCHIKEY>
><E_SAVI_BUILDING_BLOCK_B_PROTECTION_NEEDED>
><E_SAVI_BUILDING_BLOCK_B_PROTECTED>
><E_SAVI_BUILDING_BLOCK_B_ORDER_LINK>

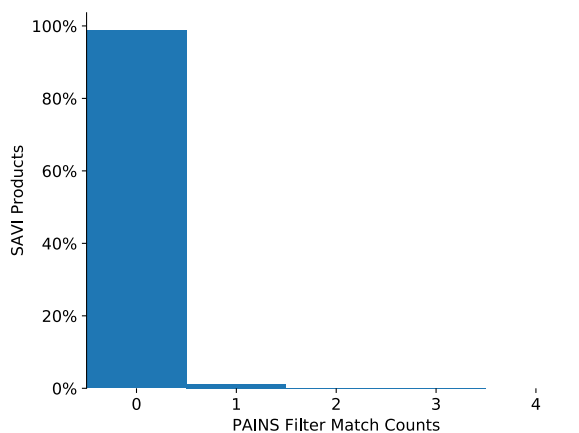
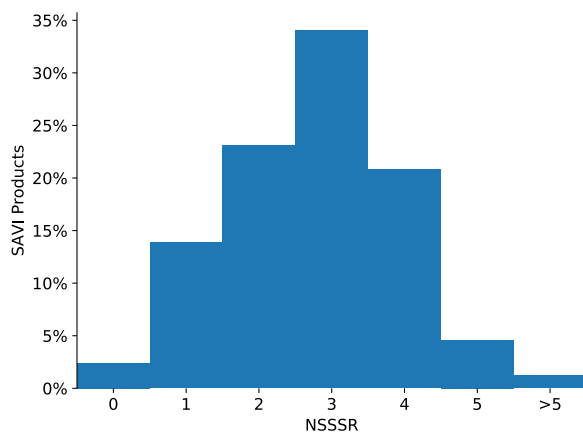
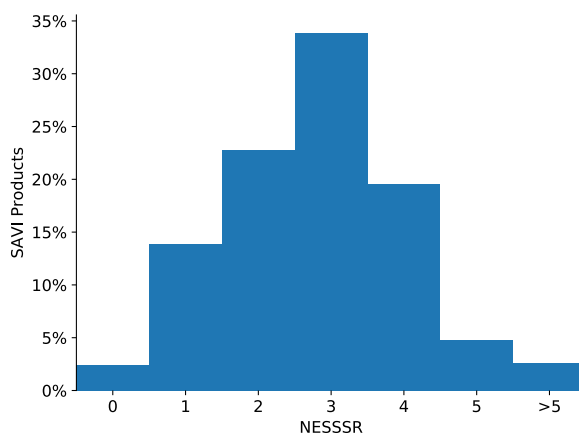
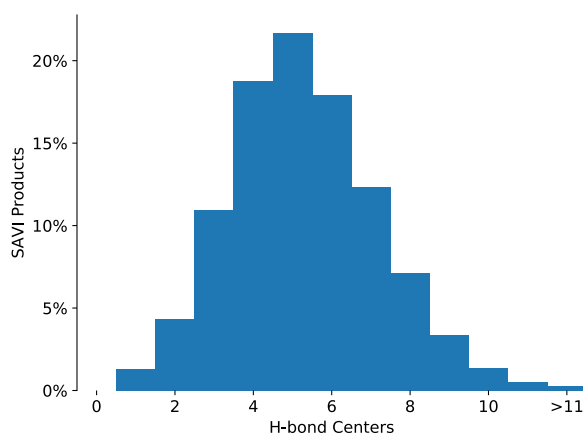
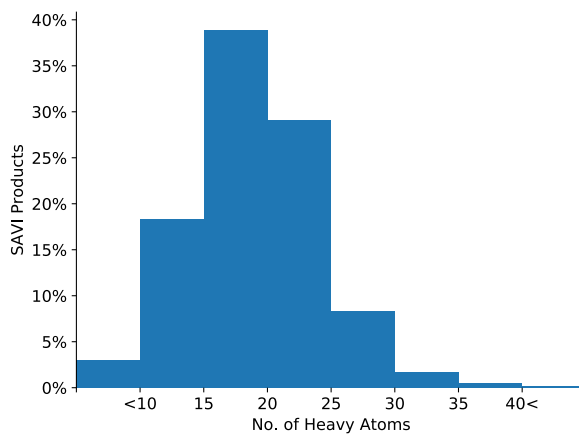
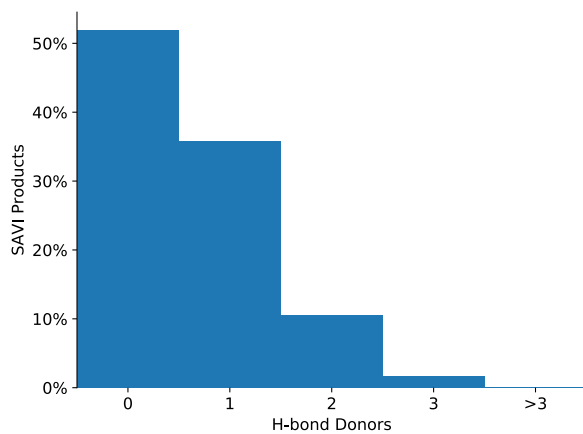
4. Additional plots of property distribution for “Plus” set

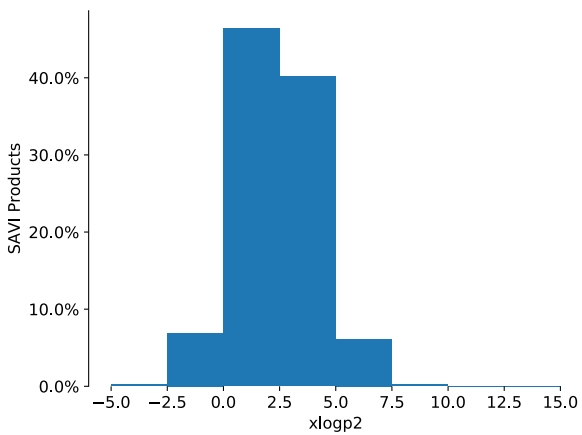
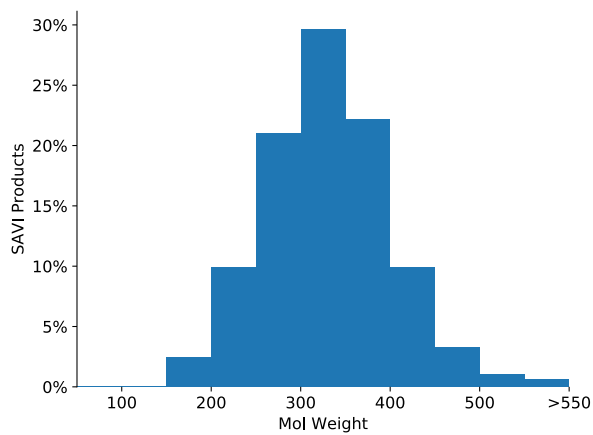
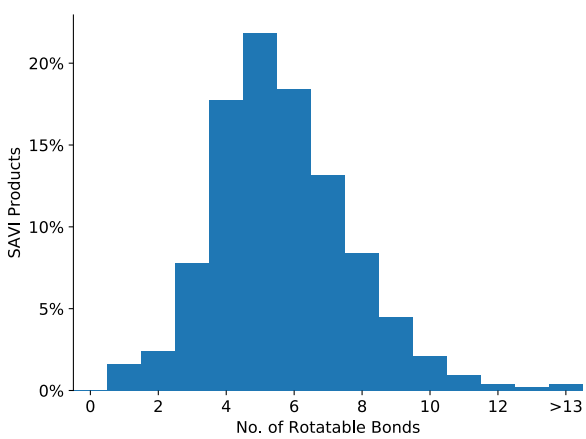
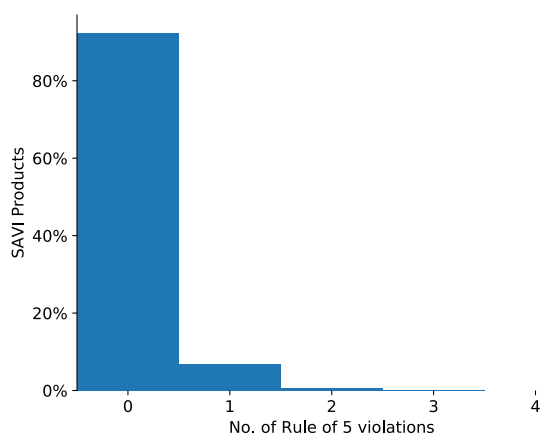
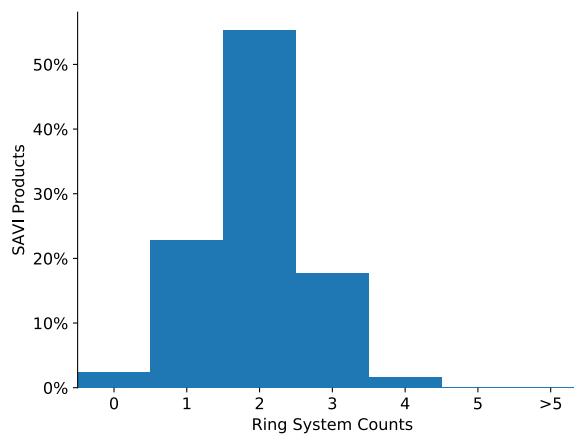
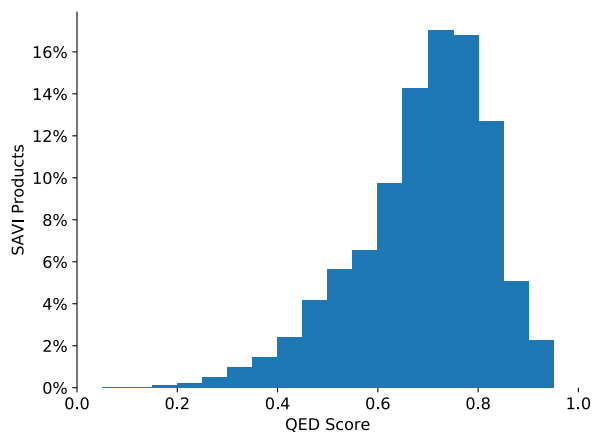




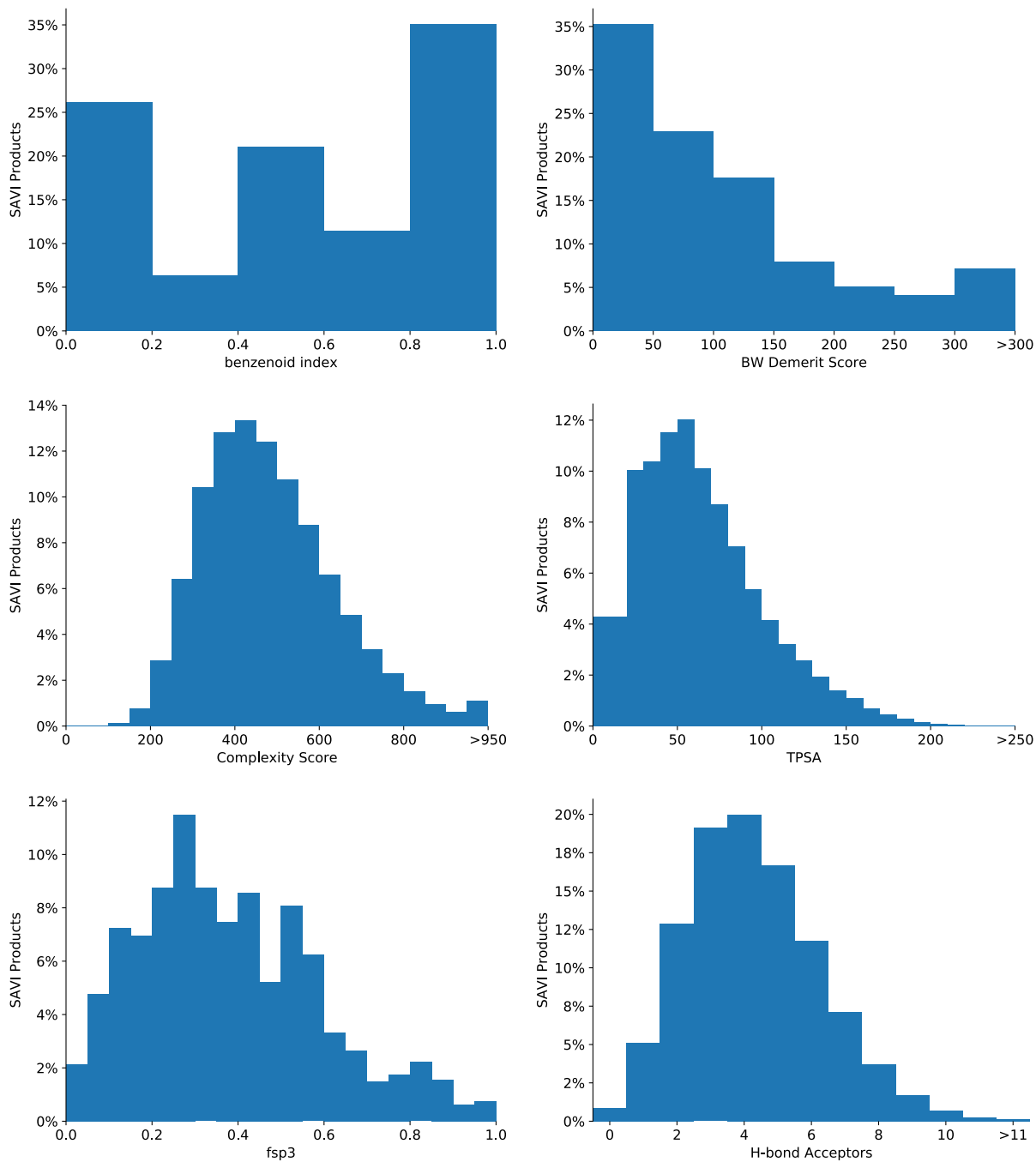
5. Plots of property distribution for “Neg0” set

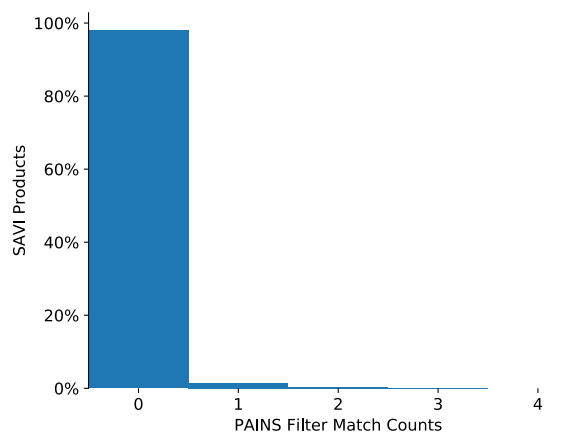
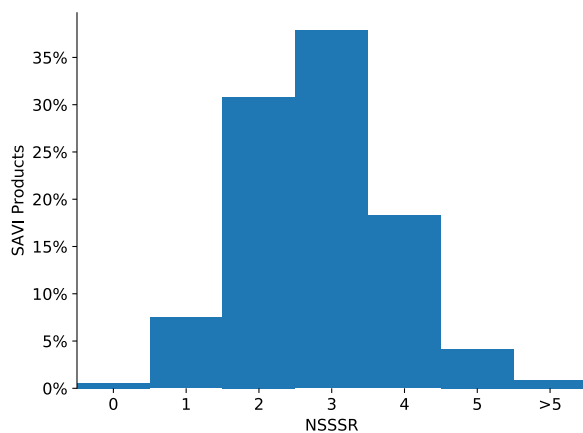
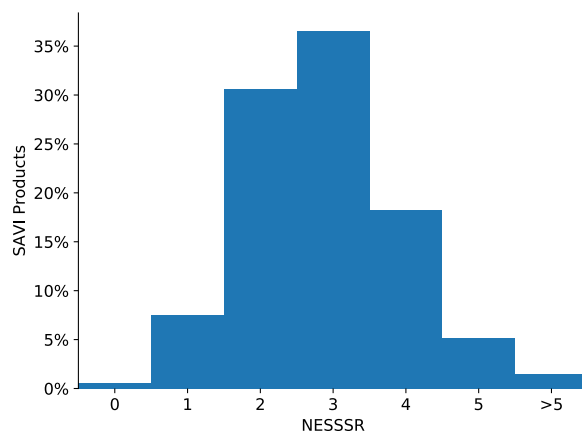
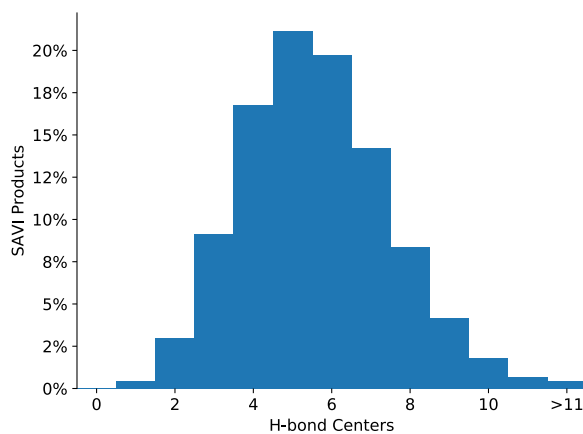
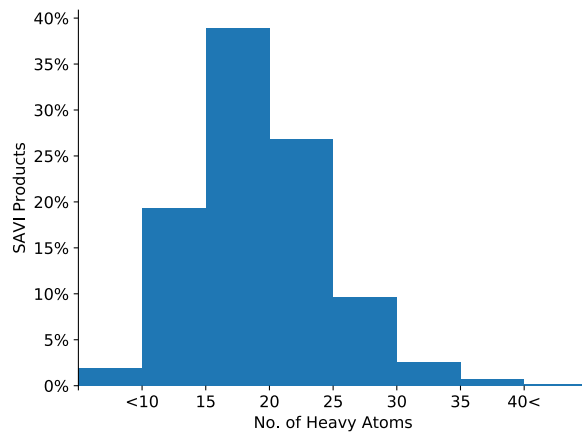
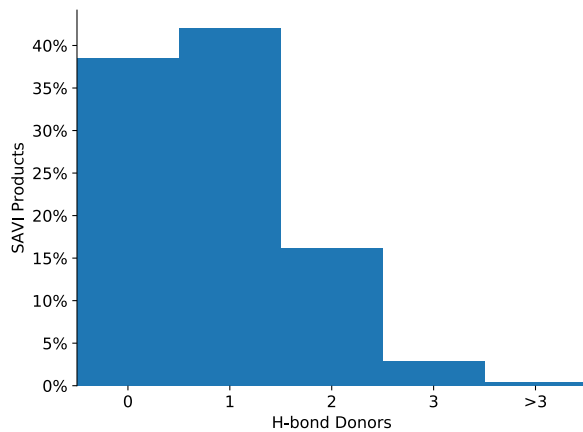


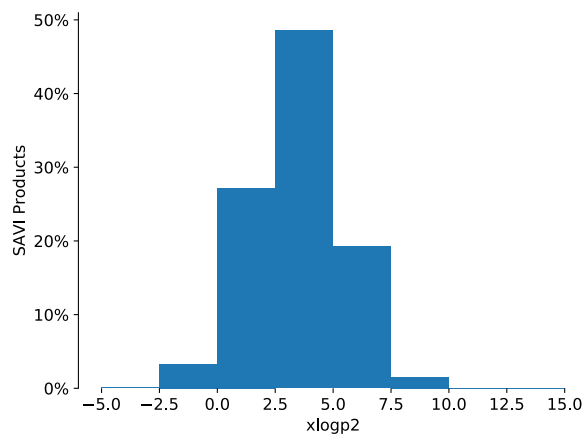
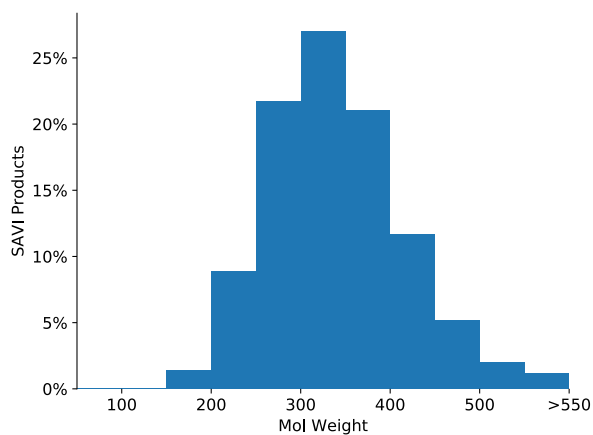
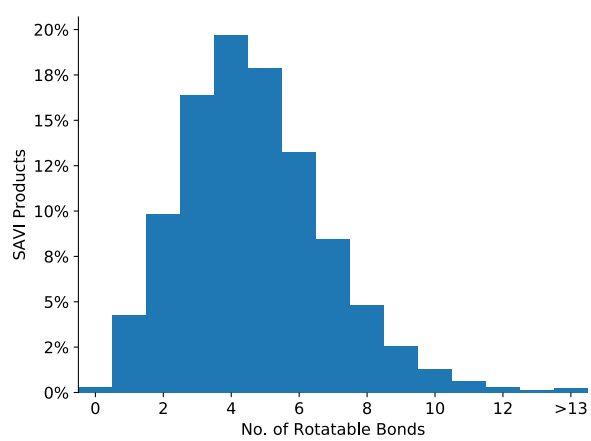
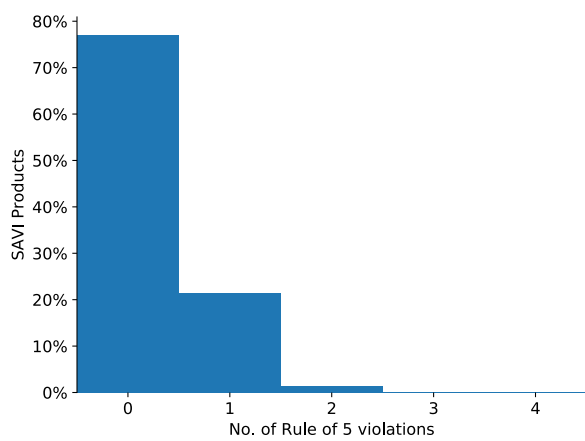
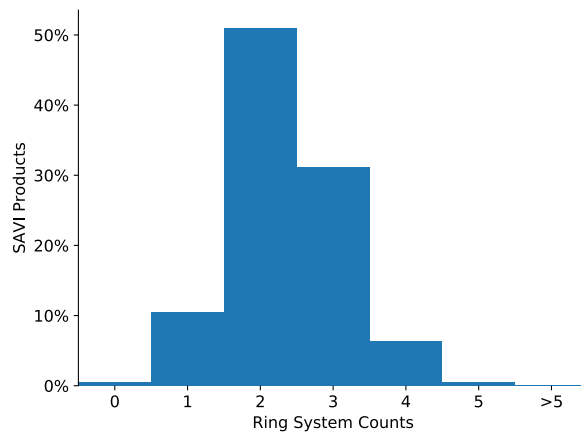
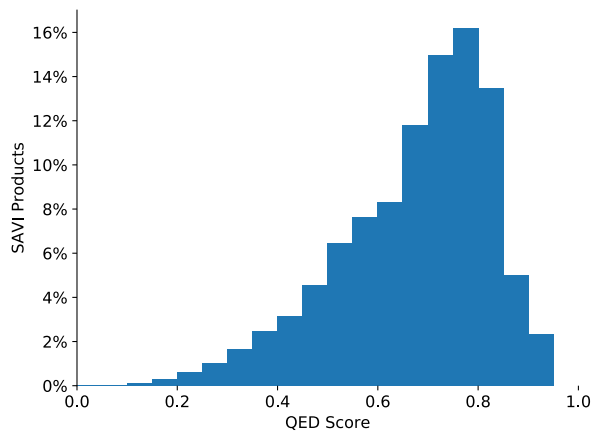




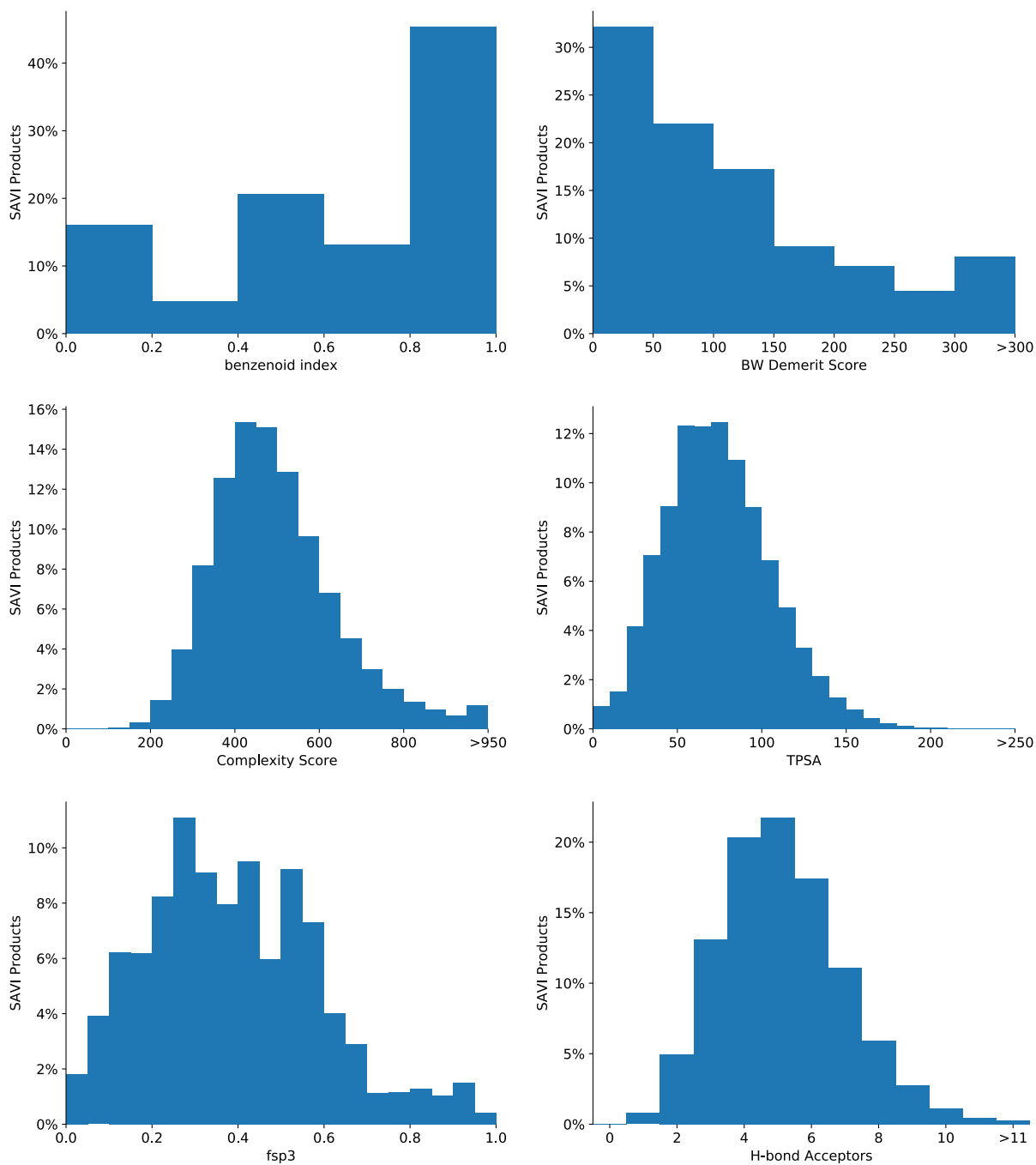
6. Plots of property distribution for “Neg10” set

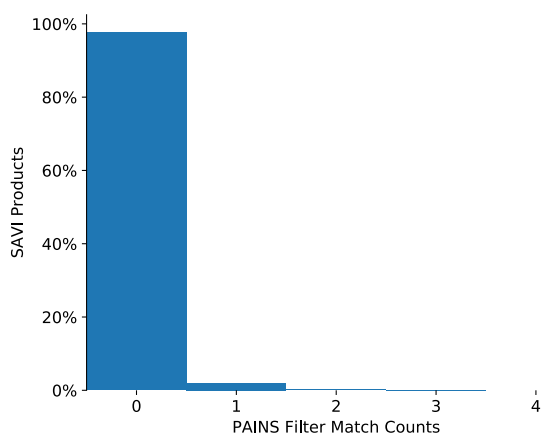
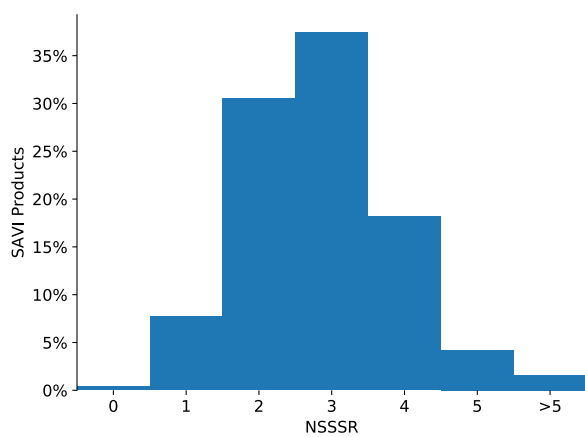
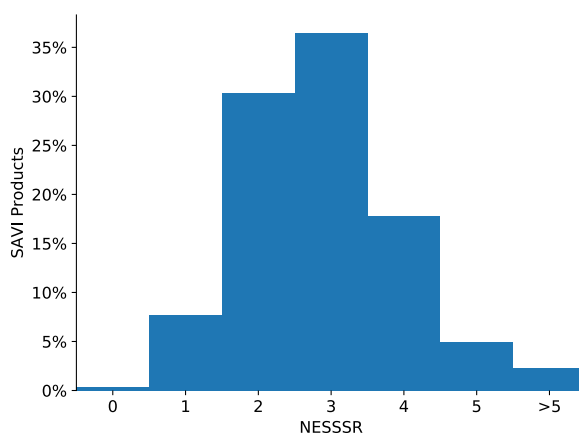
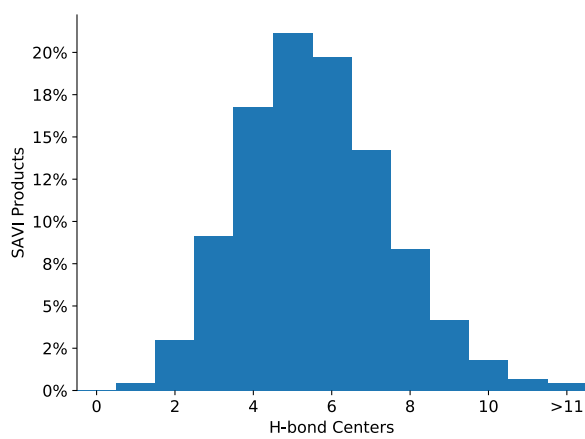
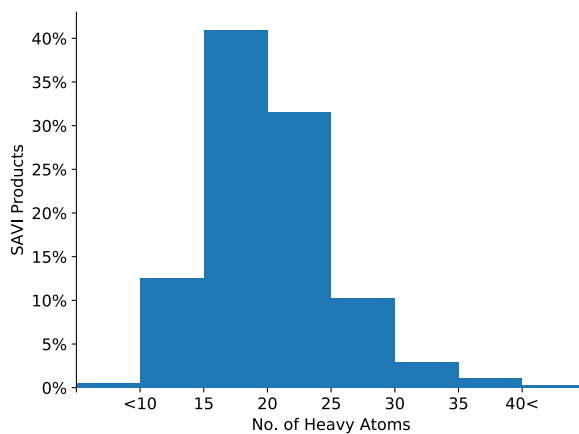
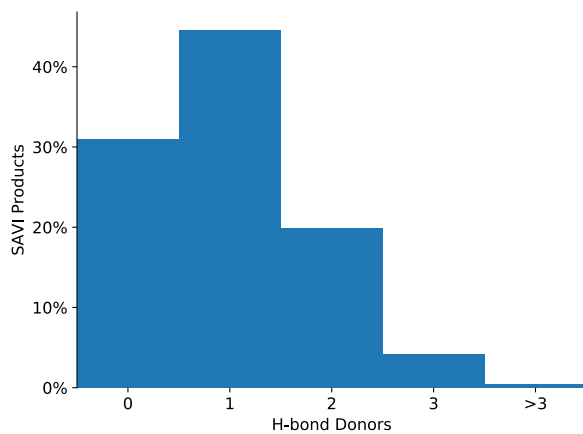


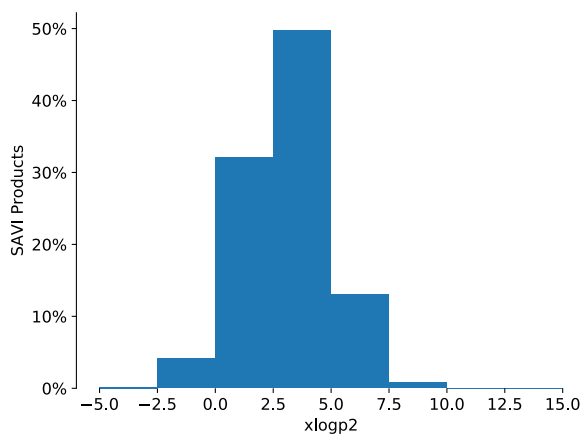
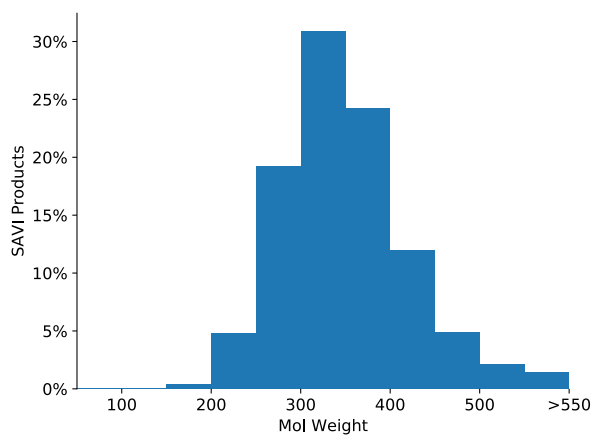
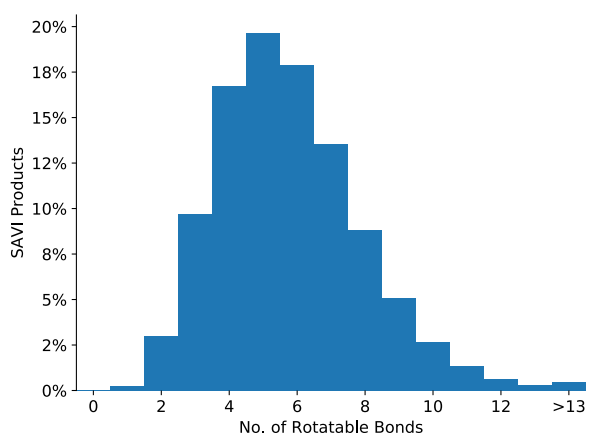
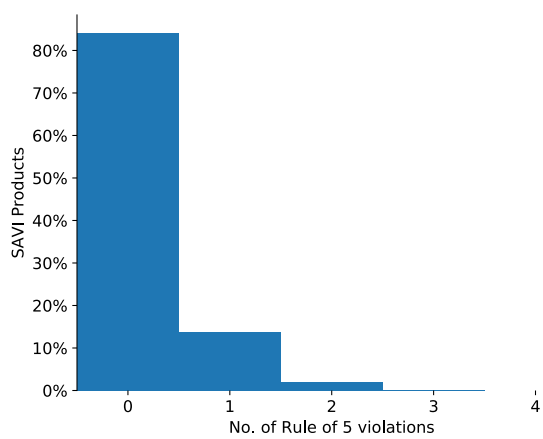
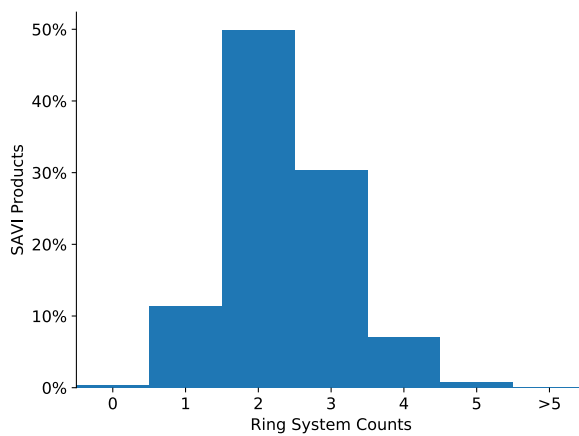
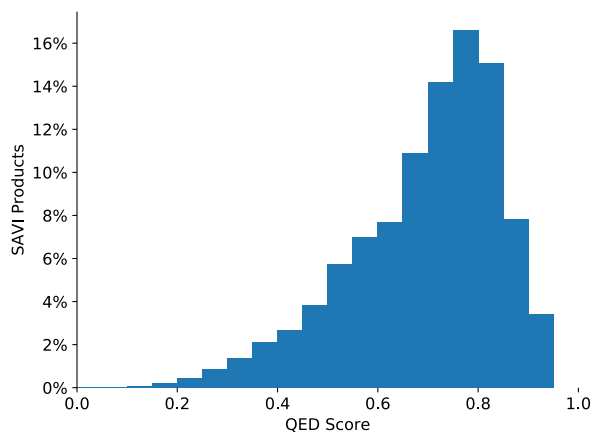




7. Plots of property distribution for “Neg20” set







8. Plots of property distribution for “Neg30” set

