Computational prediction of the molecular configuration of three-dimensional network polymers

Lies De Keer,[1] Karsu I. Kilic,[2] Paul H.M. Van Steenberge,[1] Lode Daelemans,[3] Daniel Kodura,[4] Hendrik Frisch,[4] Karen De Clerck,[3] Marie-Françoise Reyniers,[1] Christopher Barner-Kowollik,[4,5] Reinhold H. Dauskardt,[2*] Dagmar R. D'hooge[1,3*]

[1]Laboratory for Chemical Technology (LCT), Ghent University, Technologiepark 125, 9052 Ghent, Belgium

[2]Department of Materials Science and Engineering, Stanford University, 416 Escondido Mall, Stanford, CA 94305-2205, USA

[3]Centre for Textiles Science and Engineering (CTSE), Ghent University, Technologiepark 70a, 9052 Ghent, Belgium

[4]Centre for Materials Science, School of Chemistry and Physics, Queensland University of Technology (QUT), 2 George Street, Brisbane QLD, 4000, Australia

[5]Macromolecular Architectures, Institut für Technische Chemie and Polymerchemie (ITCP), Karlsruhe Institute of Technology (KIT), Engesserstrasse 18, 76128, Karlsruhe, Germany

Corresponding authors: Dagmar R. D'hooge and Reinhold H. Dauskardt

E-mail: dagmar.dhooge@ugent.be and dauskardt@stanford.edu

Abstract

The three-dimensional arrangement of natural and synthetic network materials determines their application range. Control over the real time incorporation of each building block and functional group is desired to regulate the macroscopic properties of the material from the molecular level onwards. Here we report an approach combining kinetic Monte Carlo and molecular dynamics simulations that chemically and physically predicts the interactions between building blocks in time and in space for the entire formation process of three-dimensional networks. This framework takes into account variations in inter- and intramolecular chemical reactivity, diffusivity, segmental compositions, branch/network point locations and defects. From the kinetic and three-dimensional structural information gathered, we construct structure-property relationships based on molecular descriptors such as pore size or dangling chain distribution and differentiate ideal from non-ideal structural elements. We validate such relationships by synthetizing organosilica, epoxy-amine and Diels-Alder networks with tailored properties and functions, further demonstrating the broad applicability of the platform.

An essential property for network materials is their degree of three-dimensional (3D) configurational ordering, as it determines their macroscopic properties and range of applications.[1-4] Highly ordered carbon atom (C) arrangements possess exceptional characteristics such as diamond (sp^3 hybridization; Fig. 1a) and graphite (sp^2 hybridization) which are known for their hardness and lubricating potential, respectively.[5-6] Synthetic zeolites also display an ordered molecular pore structure, providing shape selectivity.[7,8]

In contrast, natural polymers (e.g. lignin[9]) and network polymers synthesized from oil-derived or renewable (co)monomers[10-16] are characterized by a less ordered configuration. Fig. 1b for instance focuses on organosilica networks made from polycondensation for use in solar cells, separation and antimicrobial coating applications. [10-12] Ideally, we aim for complete crosslinking, hence, the formation of crosslinking points (CP's) with 4 silica-oxygen-silica crosslinks (ideal CP_4 structural elements), but inevitably end up with a mixture of CP_4 elements down to CP_1. In organic chain-growth network synthesis, which has applications such as hygiene products, tissue engineering scaffolds, and drug delivery devices,[13] the situation is even more complex. We then encounter a large number of side reactions as well as a more complex reactant structure involving a plethora of functional groups.[14-16] As shown in Fig. 1c, a highly random mixture of sp^3 and sp^2 hybridizations is obtained which is accompanied by several branch types and a broad spectrum of monomer sequences (e.g. BBBAA and AABABAAAB). Moreover, the regular A/B monomer connectivity can be perturbed by structural defects (e.g. label D in zoom of Fig. 1c) and non-ideal cyclic elements that define the molecular pores (e.g. light orange elements in Fig. 1c). A variety of molecular pore sizes eventually exists because intramolecular reactions take place that involve many monomer sequence lengths. Some sequences remain dangling and, hence, do not become network segments, as diffusional limitations constrain further reaction possibilities.

Here we fully acknowledge the broad spectrum of ideal and non-ideal structural elements during the polymer network synthesis[14,17,18] to predict macroscopic properties at the molecular level. Starting from the original building blocks we target the spatial and therefore the 3D incorporation of all functional groups, segments, and structural defects at any time point of the synthesis. This goes beyond the state-

of-the art in which experimental and theoretical methodologies have never jointly addressed synthesis time dependent phenomena such as (i) the kinetic and thermodynamic interplay of chemical reactivity and molecular diffusion, and (ii) the 3D variation of compositional and topological combinations, including structural defects. We present a generic platform for (in)organic synthesis routes with model parameters that are determined according to general guidelines.

Demonstration of the conceptual framework with organosilica

Experimental methodologies for network characterization deliver information at scales larger than the molecular one where chromatography and mass spectrometry are not automatically applicable, while analytical methods such as solid-state nuclear magnetic resonance spectrometry provide only bulk information.[17,19,20] However, despite this, average characteristics such as the overall (co)monomer content and the number of branching points can be measured.[19] In general, experimental limitations exist that make it impossible to directly measure structural deviations between and within network molecules.

Also, theoretical methodologies face many restrictions. For instance, in chemical kinetics one emphasizes an overall thus averaged synthesis time dependency for the (co)monomer/functionality incorporation during network synthesis. The temporal evolution of reactant, intermediate and product concentrations has been addressed, considering side reactions, but is often limited to the time domain before gelation. Only rarely has 2D connectivity or CP clustering in (graph) node notation been reported,[21,22] however, a 3D network configuration has not yet been calculated even at high yield. The most promising approach is a matrix-based kinetic Monte Carlo (*k*MC) to simulate the molecular build-up of individual linear or slightly branched chains (*e.g.* 10^4-10^5).[14,25,26] Once average characteristics (*e.g.* the number average chain length)[16,23] are validated with inputted individual kinetic parameters from independent techniques,[24] we can access (co)monomer sequences (e.g. AAABBA vs ABABAA).

Simplifications have also been made with force-field-based molecular dynamic (MD) simulations predicting ideal defect-free 3D network structures at high conversion but neglecting kinetic constraints. A key assumption is the free movement of atoms or atom groups until the thermodynamically equilibrated configuration is reached, typically recognizing a limited number of chemical bonds, thus

ignoring many side (e.g. intramolecular) reactions and diffusivity constraints.[10,27-32] At equilibrium, MD simulations assume a generally incorrect large real time. In contrast, a real system such as free radical based organic networks are produced on a minute to hour time scale with altering dynamic viscosity (η), leading to time and chain length dependent diffusivities. This prevents the continuous incorporation of ideal structural elements, causing local structural defect formation.[3,33]

As conceptually highlighted in Fig. 2, we overcome these simplifications by connecting chemical kinetics and physics via a generic framework to visualize the spatial kinetic growth of individual network molecules. We uniquely bridge matrix-based *k*MC and MD simulations to store at any synthesis time *t* the molecular information on the individual (segment) compositions, functional groups, bond lengths, bond and dihedral angles, and structural defects. We can visualize the 3D birth of network molecules and are able to explicitly map further inter/intramolecular interactions and chemical modifications.

Fig. 2 focuses on the synthesis of organosilica networks commencing with a sufficiently large number of tetraethylorthosilicate (TEOS) molecules. As shown in Fig. 2a, the initial TEOS molecules can be hydrolyzed multiple times (top of box) and take part in condensation reactions, forming Si-O-Si crosslinks (bottom of box). A differentiation is made between inter- and intramolecular crosslinking reactions. The intermolecular reactions lead to larger network molecules, but they deplete the linear molecules by their inclusion as dangling chains. The intramolecular reactions increase the crosslinking degree within a given network molecule. They also become more likely at higher polymer network yields[34] and lead to structural defects.[35]

Every reaction step in Fig. 2a is defined by the participating functional groups and is characterized by a chemical rate coefficient (k_{chem}). Regression to isothermal experimental data should be performed to determine the k_{chem} values, using monofunctional analogues of the monomers or limited to low crosslinking yields. In the former situation, no network formation takes place so that diffusivity issues are absent and in the latter situation linear species are so dominant that there is a negligible impact of diffusional limitations. Prior *k*MC screening is recommended to associate the dominance of specific reactions to specific synthesis time regimes. This is exemplified in part 1 of the Supplementary

Discussion, which demonstrates that the number of unreacted ethoxy groups influences the hydrolysis k_{chem} value. As summarized in Supplementary Table 1, at the end, 7 intermolecular k_{chem} values result after stepwise tuning of the short synthesis time ^{29}Si-nuclear magnetic resonance data.

Once k_{chem} values in Fig. 2a are known, the focus can be shifted to the actual network formation at longer times, thereby distinguishing between the concentration (C) variations of the monomer (TEOS), dimer and ultimately any highly crosslinked molecule. As shown in Fig. 2b, to obtain C values, inter- and intramolecular reaction rates ($r_{inter/intra}$) need to be calculated that are corrected for diffusional limitations. Intermolecular apparent rate coefficients (k_{app} instead of k_{chem} values) are introduced to determine the interaction of intrinsic kinetics and diffusional constraints.[36] For larger η, the intermolecular reactivities can even be dictated by chain length and CP dependent diffusion coefficients (D values). The parameters to calculate the intermolecular diffusion coefficients (D_{inter} values) follow from independent techniques and network scale validation based on long time data.[14] Intramolecular Si-O-Si crosslinking becomes impossible if the distance between the functional groups is too large (red vs. green arrows in Fig. 2b), since the intramolecular diffusion coefficients of the functional groups (D_{intra} values) are then too low. Therefore, fundamental distance rules addressing rigidity variations are introduced.

Such detailed molecular scale analysis is possible, as during the calculation of the $r_{inter/intra}$ values, which are a function of k_{app} and C variations, we keep track of the connectivity history for the *k*MC ensemble through a composite topology matrix T. We can retrieve extensive molecular distributed data regarding (i) the composition of the remaining and formed linear/branched molecules, and (ii) the segment compositions, including structural defects, and the connectivity of each separate network molecule (Fig. 2c; top) at any t. Combined with MD simulation input, a 3D representation of all (network) molecules is within reach. MD simulations allow access to a spectrum of bond and dihedral angles, and bond lengths that are thermodynamically feasible,[10] as illustrated in Fig. 2c (bottom). By following each connectivity in T the local environment is scanned, and a feasible 3D structural element is generated if no structural defect is detected. Upon doing so, alternative 3D visualization rules are needed to ensure the correct representation of the non-ideality. MD simulations deliver input for the *k*MC algorithm and,

hence, any development, e.g. the improved description of polymer-medium interactions, can be directly translated in the current multi-scale development.

As highlighted in Fig. 2d-e, the 3D molecular information from Fig. 2a-c can be utilized to construct fundamental structure-property relationships at any *t*. This *t* dependency constitutes the strength of our framework, as macroscopic property design is achieved whenever desired, and not restricted to simplified cases of limiting crosslinking or ideally constructed high yield (theoretical) networks. As soon as (i) chemical model parameters have been determined under conditions with no kinetic impact of viscosity variations and (ii) diffusion model parameters have been validated based on longer time data, our framework can be applied.

From molecular descriptors to organosilica properties

Fig. 3 applies the concepts outlined in Fig. 2 to understand organosilica network synthesis undertaken at 298 K, considering initial water to TEOS molar ratios between 0.3 and 10. The parameters are listed in Supplementary Table 1 and their reliable determination is confirmed in Supplementary Figure 8 and 9.[37] Part of this validation is repeated in Fig. 3a, depicting an excellent agreement between Si-O-Si experimental and simulated yield for an initial water to TEOS molar ratio of 4. We include a 3D visualization at three timestamps, while at t_3 a simplified 2D visualization is also shown.

Fig. 3b gives two examples of molecular scale insight by tracking *t* variations in *T*. We plot (i) the concentration of network molecules with more than 3 CP's possessing at least 3 crosslinks (CP_3+CP_4; left axis; black) and (ii) the fraction of intramolecular reactions (f_{intra}; right axis; green). We display in Fig. 3c-e distributed molecular descriptors at the 3 timestamps. We show the (i) number (CP_3+CP_4) distribution (Fig. 3c), (ii) mass chain length distribution (Fig. 3d), and (iii) number molecular pore size distribution (Fig. 3e). In Fig. 3c, an inset is included for the (CP_3+CP_4) distribution showing theoretical (ideal) results for f_{intra}=0 (no intramolecular reactions), also depicting the ideal 3D representation at t_3. For the chain length distribution at t_2 the contribution of the sol and the gel (blue/green symbols) is differentiated in the inset in Fig. 3d.

We see in Fig. 3a-e that a gradual transition from a linear to a slightly crosslinked polymer is obtained that is fully converted into a network material. This is evidenced by the increasing concentration of network molecules in Fig. 3b with at least three CP's crosslinked three times. A high degree of structural heterogeneity is always established, as obvious from the shape time dependencies in Fig. 3c-e. At short times (t_1) the (CP_3+CP_4) distribution in Fig. 3c is less-defined, particularly for molecules with less crosslinking, whereas at longer times (t_2 and t_3) the non-ideality is also established at the tail. This non-ideality is even clearer in the insets, displaying for f_{intra}=0 strong deviations due to a completely different (ideal) 3D representation. Tail dependencies are also valid for the chain length distribution in Fig. 3d. The evolution of the molecular pore size distribution in Fig. 3e is also complex. First a bimodal character is manifested (t_1 to t_2), as f_{intra} increases from 0 to a significant value of 0.3 and more pores are formed in the heterogenous molecules with respect to their size. At longer time (t_3), this bimodality disappears as a substantial share of intramolecular reactions, which implies abundant small pore formation, shifts the distribution to the left making it monomodal. The number average pore size thus displays a maximum at an intermediate polymer yield. This is further confirmed in Supplementary Figure 14, also depicting non-trivial temporal variations for molecular properties such as the CP distance distribution, and the average crosslinking density and number/mass chain length.

Changing the reaction conditions further complicates the competition between inter- and intramolecular reactions and their interplay with diffusional limitations. This affects the pattern in which functional groups are positioned, as exemplified by a variation of simulated hydrophilicities in Fig. 3f (left axis; black symbols) achieved by fixing the Si-O-Si yield at 70% but altering the initial molar ratio of H_2O to TEOS. Post-processing[39] allows calculation of the number of OH groups near the surface so that the hydrophilicity can be simulated via normalization with respect to the total number of functional groups. We observe in Fig. 3f (right axis; turquoise symbols[40]) an inverse correlation with the contact angle. Upon closer inspection this angle is only sensitive in the region of lower to intermediate hydrophilicities, as associated with initial water to TEOS molar ratios between 0.3 and 4. The modeling results show that higher initial molar ratios are worthwhile if one wants to maximize OH groups near/on the surface. Similarly, very low hydrophilicity results for very low molar ratios. In a broader perspective, such an

insight is crucial for instance for molecular scale driven separation,[12,41] as illustrated in Fig. 3g (right axis). To retrieve a highly pure water filtrate out of a water/toluene azeotropic mixture, dedicated hydrophilicity design is required. Upon applying stronger hydrolysis conditions which result in higher surface hydrophilicity, the filtrate is more toluene-rich. The water molecules stay on top of the membrane due to stronger interactions, consistent with the higher water adsorption isothermal data in Fig. 3g (left axis). Only upon quantitatively converting ethoxy into OH groups and condensation to Si-O-Si, the surface hydrophilicity can be minimized so that the lowest adsorption data are recorded and the capillary effect can induce a water-rich filtrate.

Going beyond the organosilica case

For an epoxy-amine curing case shown in Fig. 4a-b, as relevant for coatings,[38] emphasis is placed on the *t* dependence of the storage modulus (Fig. 4b; right axis; turquoise). The initial (stoichiometric) conditions and parameters are listed in Supplementary Table 2. The reliability follows from accurately describing spectroscopic data in Supplementary Figure 20a.

Our platform reveals that two molecular descriptors matter with a different *t* window. The first descriptor is the concentration of molecules with x=3 CP's containing at least 3 crosslinks (CP_3+CP_4) and the second is f_{intra}. These descriptors are included in Fig. 4b as a black dashed dotted line (left axis) and a black full line in the inset. At short times, the material stiffness is limited with a steadily increasing but always low storage modulus. To significantly increase the modulus, we first need a substantial number of molecules with 3 CP's of the type CP_3 and CP_4. Extra rigidity/stiffness is gradually acquired at longer synthesis times, at which point the contribution of molecules with more than 3 CP's of the type CP_3 and CP_4 becomes gradually more relevant (other dashed dotted lines in Fig. 4b; left axis; x=4-6). This gradual increase is too slow to explain the steep increase in modulus at very long times, consistent with the much slower increase of the conventional average crosslinking density in Supplementary Figure 20e. The only descriptor that follows this steep increase is f_{intra}. Hence, again the non-ideal structural elements matter, which is in agreement with the significant cyclization thus loop formation in the 3D ensemble of epoxy-amine network molecules as shown in Fig. 4a. As shown in Supplementary Figure 20b, under non-stoichiometric conditions, fewer loops are predicted at the final yield, indicating lower strength. Indeed,

as shown in the same subplot, the experimental storage modulus decreases, specifically if the concentration of network molecules with 3 CP's of type CP_3 and CP_4 and the values for f_{intra} drop.

In Fig. 4c-d, we further employ in silico derived molecular scale information to understand variations in swelling degree, a key property for hydrogels and drug delivery.[42] We apply Diels-Alder network chemistry[43,44] at 353 K and focus on the final yield, selecting tetrafunctional and bifunctional building blocks. The initial conditions and parameters are provided in Supplementary Table 3, with validation provided in Supplementary Figure 14. The swelling degree (Fig. 4d; right axis; turquoise) refers to the relative amount of incorporated toluene compared to the dry gel. The CP, the dangling chain and the molecular pore size distribution from Supplementary Figure 22a-c are utilized to calculate the average molecular descriptors (Fig. 4d; left axis; black and blue) and to understand the swelling data. The further away we are from the stoichiometric situation, as defined by an initial molar ratio of bi- to tetrafunctional linker (r) of 2, the more swelling is observed, as shown in Fig. 4d (right axis; turquoise symbols). The simulated average number of CP_x ($x \geq 3$; left axis; black symbols) then lowers and the network structure is more open. Also, more dangling chains remain and the simulated number average molecular pore size increases (left axis; blue symbols), explaining the larger swelling degree.

To ensure less swelling we need to synthesize more compact high yield 3D Diels-Alder networks. This is illustrated in Fig. 4c, addressing high yield network molecules for r=2 which display the highest f_{intra}. The OH groups associated with bifunctional monomeric building blocks are marked in purple, whereas all other moieties remain black. These OH groups are a handle for chemical modification (e.g. drug loading) but one cannot simply assume a pre-defined ideal functional group distance distribution for their high yield incorporation.

For the idealized case of f_{intra}=0 in Fig. 5a (purple symbols; r=2; additional results in Supplementary Figure 22d-f), a broad OH group distance distribution results, representing a wide variety of network molecules. In reality, intramolecular reactions lead to significant loop formation so that the OH group distance distribution becomes bimodal and is shifted to the left (full purple symbols in Fig. 5a). As illustrated in Fig. 5b-c, this effect of f_{intra} on the OH group distance distribution affects the material performance, in this case the tracer release potential. In Fig. 5b, 50% of the OH groups (70% yield) are

chemically modified into the Fmoc tracer (purple to orange), under equimolar and non-equimolar conditions (*r* of 2, 1.5 and 1). Upon contact with 1,8-diazabicyclo[5.4.0]undec-7-ene (DBU), Fmoc is released and from the three bottom curves in Fig. 5c, it follows that the slowest release results for *r*=2.

The release order at the bottom of Fig. 5c can be explained by comparing the average pore size (PS_{av}) data. Due to the high f_{intra} for *r*=2, a 3D network results with a low average pore size (PS_{av}) of 16 units. Upon considering for example the non-equimolar case of *r* equal to 1.5, PS_{av} already increases to 29 units. This is consistent with the much more open structure in Fig. 5b for *r*=1.5 compared to *r*=2 (top vs. bottom box), facilitating DBU and Fmoc diffusion. The molecular descriptor PS_{av} from the explicit calculation of the molecular pore size distribution can therefore be correlated with the experimentally accessible release rate coefficient $k_{release}$ (s^{-1}). This is highlighted in Fig. 5c considering PS_{av} data of high yield networks at *r*=1, 1.5 and 2. We select an exponential function, indicating that a single cyclic molecule with a maximum pore size should release instantaneously. This physical boundary is supported by the very high $k_{release}$ value (> 0.01 s^{-1}) under initial conditions leading to quasi-linear (and star) oligomeric species even at maximum yield (*r*=2.5; excess of bifunctional monomer). Moreover, upon lowering *t* (*r*=2) but still ensuring sufficient network formation, our platform predicts a higher PS_{av}, which is translated into a much faster release process as depicted by the top curve in Fig. 5c. Therefore, we cannot rely on correlations based simply on an *r* variation, as the key intrinsic network material characteristic is PS_{av}, which is dependent on both the initial conditions and the final synthesis time.

Outlook

Fig. 3-5 highlight that our framework provides a detailed characterization and design of (in)organic polymer network materials from the molecular scale right up to the application level. We can evaluate if a network material is close to the desired target structure, and can run the framework over a wide range of conditions to reduce the amount of costly experiments to improve material performance. Non-idealities and 3D molecule-to-molecule deviations matter strongly and the molecular scale interpretation relies on the calculation of average molecular descriptors explicitly acknowledging such distributed deviations.

Our work is also relevant for emerging fields that exploit network materials. A first example consists of network materials derived from supramolecular chemistry, in which guest-host interactions of only the well-positioned functional groups are expected to lead to advanced sensing and recognition.[45,46] Another example consists of dynamic recyclable networks in which reshaping can be accelerated if one knows the exact 3D position of the exchangeable groups.[47,48]

Figure captions

Fig. 1: Network materials with decreasing order of their 3D structural configuration. **a** perfectly ordered carbon (C) atoms with sp^3 hybridization, as in diamond. **b** highly ordered silica atoms with oxygen (O) bridges with sp^3 hybridization with a limited number of structural defects, as in solar cells. Examples of defects are unreacted functional groups (FG^1 = OEt, FG^2 = OH; Et= ethyl; OH= hydroxyl), leading to incomplete crosslinking points (CP's; CP_i = CP with *i* crosslinks; 1 to 4). **c** highly random configurations of (co)monomer units (mixture of sp^3 and sp^2 hybridization) containing atoms such as C, O and nitrogen (N), and functional groups leading to network molecules with different segment lengths and many structural defects (*e.g.* dangling chains: short and long branches (S/LCB's), and inverse monomer insertions), as in hydrogels; examples of molecular pores (mP's) are colored in **c** in light orange; the building blocks are symbolized by colored spheres that represent atoms (in **a** and **b**) and a chemical moiety such as a targeted comonomer unit A/B or side reaction defect D (in **c**).

Fig. 2: Concepts of the framework exemplified for organosilica synthesis using tetraethylorthosilicate (TEOS); matrix-based kinetic Monte Carlo (*k*MC) and molecular dynamic (MD) simulations are interconnected to visualize the incorporation of building block by building block so each functional group (FG) at any synthesis time *t*; illustration. **a** molecular scale with chemical rate coefficients for main and side reactions (k_{chem} values; Arrhenius parameters A and E_a; reaction distance σ; polymerization temperature T_p and pressure p_p). **b** chemical kinetics with calculation of inter/intramolecular rates ($r_{inter/intra}$; (mol L^{-1}) s^{-1}) from concentration of reactants, intermediates and products (C values) accounting for interplay of inter/intramolecular reactions and diffusivities (D: diffusion coefficient; η: dynamic viscosity; k_{diff}: diffusion rate coefficient). **c** *k*MC data storage and update of compositions, including segments, structural defects, and connectivity of crosslinking points (CP's). Combined with MD data on bond/dihedral angles and bond lengths 3D configurations for each molecule result at any *t* accounting for non-idealities and thus, defects. **d** in silico derived molecular network characteristics up to topological scale at any *t*; examples are the CP distribution (at least 3 crosslinks; CP_3+CP_4), chain length distribution, and molecular pore size distribution. **e** characteristics from d serve as input to construct structure-property relationships that evaluate the network material performance, with a possible differentiation between ideal and non-ideal structural network elements. Example of hydrophilicity calculation based on the presence of OH functional groups near the surface (see Supplementary Methods).

Fig. 3: Framework application for organosilica network synthesis with tetraethylorthosilicate (TEOS); **a** yield (Supplementary Equation 5; experimental data from ref [37]) as a function of synthesis time *t* and birth/3D growth of network molecules in 3 boxes; initial conditions and model parameters: Supplementary Table 1; model validation: Supplementary Figure 8 and 9; FG: functional group; **b** concentration variation of molecules with > 3 crosslinking points (CP's; CP_3+CP_4) alongside the (cumulative) fraction of intramolecular contributions, f_{intra}; **c**-**e** distributed molecular descriptors for three synthesis times (t_1, t_2 and t_3), with in subplot **c** number CP distributions (CP_3 + CP_4) with in its inset the results for the idealized case (f_{intra}=0), also including a 3D representation (t_3); in subplot **d** mass chain length distributions (CLDs) where a monomer unit represents an incorporated initial building block and the inset distinguishes at t_2 between sol (linear/loosely branched chains) and gel (precipitated molecules) CLDs, considering as gel cut-off number of CP's to number of monomer units ratios larger than 0.5; in subplot **e** number molecular pore size distributions (pore size: number of monomer units in the pore); **f** Structure-property relationship at the molecular level: simulated hydrophilicity at 70% yield (density-based outlier detection:[39] relative OH contribution near surface) versus initial water (H_2O) to TEOS molar ratio to understand contact angles[40] (turquoise symbols; right axis); **g** separation of water/toluene azeotropic mixture through TEOS-based membranes: need of hydrophilicity surface control for water-rich filtrate; experimental data on water adsorption isotherms (50% humidity; turquoise symbols; left axis) and filtration efficiency (yellow symbols; right axis) from ref 12; 10% error bars; n=3.

Fig. 4: Structure-property relationships for two other network chemistries other than the organosilica case. **a-b** application for epoxy-amine curing (network chemistry 2; initial and model parameters: Supplementary Table 2) with in **a** 3D visualization of representative network molecules at synthesis end and in **b** explanation of temporal variation of experimentally recorded variation of storage modulus (turquoise; right axis; this work; initial storage modulus due to the flexible cuvette) based on simulated concentration of network molecules with more than *x* (3-6) crosslinking points (CP's) containing at least three crosslinks (left axis; CP_3 + CP_4) and relative importance of intramolecular reactions (upper graph; f_{intra}); **c-d** application for Diels-Alder based case (network chemistry 3; initial and model parameters: Supplementary Table 3) with in **c** 3D visualization of representative network molecules at synthesis end (stoichiometric case) and in **d** relation between stoichiometry for initial building blocks for synthesis at 353 K and simulated (i) number average of CP's with at least 3 crosslinks (black; left axis) and (ii) number average molecular pore size (blue; left axis) at final yield to understand experimental swelling data (turquoise; right axis; this work; toluene; Supplementary Figure 27a-b); 10% error bar; n=3.

Fig. 5: Relevance of non-idealities for three-dimensional (3D) Diels-Alder based network structures (chemistry 3). **a** high yield (70%) thus long synthesis time 3D representation from Fig. 4c (molar ratio *r* of 2) compared with its ideal counterpart with no intramolecular defects (f_{intra}=0), showing a more complex variation of the OH group distance distribution for the former (full vs. open purple symbols). **b** 3D structure if 50% of the high yield OH groups are replaced by the drug-related tracer Fmoc (Methods section; chemical modification of OH bifunctional linker by reaction with succinic anhydride **1** and N-Fmoc-1,3-propanediamine hydrochloride **2**) for *r* equal to 2 and 1.5 alongside the molecular pore size distributions, also including the result for a smaller time (low yield; still sufficient network formation) at *r* equal to 2. **c** Corresponding Fmoc release curves upon addition of diazabicyclo[5.4.0]undec-7-ene (DBU) considering experimental data in Supplementary Figure 27 (quantification of 9-methylene-9H-fluorene **3**). Simulated lines are based on network release rate coefficients ($k_{release}$ values) that are correlated with the network average molecular pore size (PS_{av}), as explicitly calculated from the complete distribution and for maximal PS_{av} a very high $k_{release}$ consistent with the very fast release for quasi-linear thus not network molecules (e.g. *r*=2.5; high yield; experimental data also in Supplementary Figure 27); 10% error bar; n=3.

Methods

Part A: Experimental framework

For chemistry 1, tetraethylorthosilicate (TEOS; **4**) is used as the organosilica precursor, considering literature data[37] at 298 K on both the synthesis and the characterization. The initial mixture consists of 0.022 L TEOS, 0.0072 L H_2O (pH=2.5), and 0.022 L ethanol (EtOH) in a molar ratio l:4:3.8 with hydrogen chloride (HCl) as catalyst. By selecting a pH of 2.5 near the isoelectric point of the silica sol, hydrolysis and condensation steps can be partially separated and studied by ^{29}Si nuclear magnetic resonance (NMR), facilitating parameter tuning of the associated k_{chem} values. The NMR spectra were recorded by Pouxviel et al.[37] at 298 K and 80 MHz employing an AM400 wide bore Bruker spectrometer. The acquisition time was 0.5 s and a relaxation delay of 5 s was used. Exponential multiplication was applied before Fourier transformation. The experimental data are provided as symbols in Fig. 3 and Supplementary Figure 8 and 9.

For chemistry 2, bisphenol-F diglycidyl ether (DGEBF; **5**; Sigma-Aldrich, used as received) and ethylene diamine (EDA; **6**; Sigma-Aldrich, used as received) were used as precursors to perform the epoxy-amine curing at 298 K. The curing was analyzed with near-infrared (nIR) spectroscopy employing an in-house Perkin Elmer Lambda 900 UV/VIS/NIR Spectrometer. Both the DGEBF and EDA precursor were thoroughly mixed first under equimolar functional group concentrations and later under non-equimolar conditions, before being brought into a nIR-transparent cuvette. nIR spectra of the reaction mixture were subsequently recorded at frequent time intervals from 4000 to 10000 cm^{-1} with a resolution of 3 nm. The concentration of primary, secondary and tertiary amine, epoxide and hydroxyl groups were analyzed as a function of synthesis time using Spectragryph software, according to the procedure previously reported.[49] In this procedure, the spectra are internally normalized by the aromatic ring peak at 4620 cm^{-1} and then the primary amine and epoxide concentration are determined by integration of the peaks at 4925 and 4520 cm^{-1}. Based on these concentration profiles, the three major reactions from Supplementary Figure 16 could be followed upon using a mass balance approach[49] with the kinetic results given in Supplementary Figure 20a.

The mechanical response variation more specifically the variation of the (storage) modulus as a function of synthesis time, as needed for the construction of Fig. 4b, was recorded using a Dynamic Mechanical Analyzer (DMA Q800, TA Instruments) at room temperature, considering a precursor mixture with equimolar functional groups as brought in a flexible cuvette that was tested under three point bending with following settings: 1 Hz, 30 µm amplitude, and 15 mm span. Due to the flexible cuvette necessary to carry the reaction mixture during DMA analysis, an initial storage modulus of 20 MPa is obtained for the unreacted resin (see double turquoise arrow in Fig. 4b). As the sample holder remains the same during the experiment, there is no further influence on the curve trend.

For chemistry 3, a small molecule experiment (Supplementary Figure 21; phase 3a) with monofunctional reactants, representing the functional groups of the multifunctional monomeric building blocks for the actual Diels-Alder network formation (Supplementary Figure 17), was first conducted. Phase 3b consists of the actual network formation (use of only the OH-based bifunctional linker) and Phase 3c to a similar network formation but with drug release capacities (partial use of Fmoc-modified bifunctional linker.

For the three phases 3a-3c, ^{1}H- and ^{13}C-NMR spectra were recorded on a Bruker System 600 Ascend LH, equipped with a BBO-Probe (5 mm) with z-gradient (^{1}H: 600.13 MHz; ^{13}C: 150.90 MHz,). All measurements were carried out in deuterated solvents (Supplementary Figure 25). Resonances were recorded in parts per million (ppm) relative to tetramethylsilane (TMS). The δ-scale was calibrated to the respective residual solvent signal. The measured coupling constants were calculated in Hz. To analyze the spectra the software MESTRENOVA 11.0 was used. The signals were quoted as s = singlet, bs = broad singlet, d = doublet, t = triplet, dd = doublet of doublets, and m = multiplet.

Liquid chromatography – electrospray ionization – high resolution mass spectroscopy (LC-ESI-HRMS) measurements (Supplementary Figure 26) were additionally performed considering an UltiMate 3000 UHPLC system (Dionex, Sunnyvale, CA, USA) consisting of a pump (LPG 3400SZ, autosampler WPS 3000TSL) and a temperature-controlled column department (TCC 3000). Separation was performed on a C18 HPLC-column (Phenomenex Luna 5µm, 100 Å, 250 × 2.0 mm) operating at 313 K. A gradient of acetonitrile:water of 10:90 to 80:20 v/v (additive 10 mmol L^{-1} ammonium acetate in acetonitrile; Thermo Fisher Scientific; after drying and purification with the SP-1 Stand Alone Solvent Purification System

LC Technology Solutions Inc) was used at a flow rate of 0.20 mL min^{-1} during 15 min as the eluting solvent. The flow was split in a 9:1 ratio with 90% (0.18 mL min^{-1}) of the eluent directed through the UV-detector (VWD 3400, Dionex, detector wavelengths 215, 254, 280, and 360 nm), and 10 % (0.02 mL min^{-1}) infused into the electrospray source. Spectra were recorded on a LTQ Orbitrap Elite mass spectrometer (Thermo Fisher Scientific, San Jose, CA, USA) equipped with a HESI II probe. The instrument was calibrated in the m/z range 74-1822 using premixed calibration solutions (Thermo Scientific). A constant spray voltage of 3.5 kV, and a dimensionless sheath gas and a dimensionless auxiliary gas flow rate of 5 and 2 were applied. The capillary temperature was set to 573 K, the S-lens RF level to 68, and the aux gas heater temperature to 598 K.

Detailed spectral data are provided in Supplementary Information (Supplementary Discussion on epoxy-amine and Diels-Alder case).

Trans, trans-hexa-2,4-dienol (sorbic alcohol; **8**; reactant 1 in small molecule experiment; phase 3a) was synthesized by adding sorbic acid (**7**; 1.00 g, 8.92 mmol, 1 eq.; Sigma-Aldrich) dissolved in 50 mL diethyl ether (Et_2O; Thermo Fisher Scientific) dropwise to a 250 mL round bottom flask with lithium aluminium hydride (LiAlH; 0.94 g, 26.76 mmol, 3 eq.; 95%; Sigma-Aldrich) and 25 mL diethyl ether (Supplementary Figure 23; reaction (1)). A reflux condenser was installed and the aperture was purged with argon. After addition, the reaction was stirred at ambient temperature for 15 min and then refluxed for 1 h. The residual lithium aluminium hydride was subsequently quenched with the addition of isopropanol (Thermo Fisher Scientific), then water, following by vacuum filtration. The liquid was subsequently extracted twice with water (20 mL), dried over magnesium sulphate and the solvent removed under reduced pressure. Flash column chromatography (10:90 to 30:70 ethyl acetate:hexane) yielded the pure product as a colorless oil which crystalized over time (Supplementary Figure 24a).

Methyl 4-((((diethoxyphosphoryl)carbonothioyl)-thio)methyl)benzoate (RAFT agent; **11**; reactant 2 in small molecule experiment; phase 3a) was prepared according to the following procedure. In a two-necked 50 mL round bottom flask, equipped with a reflux condenser and a magnetic stirrer, sodium hydride (0.235 g, 9.78 mmol, 1 eq.; 60% suspension in paraffin oil; Sigma-Aldrich) and dry tetrahydrofurane (THF; 5 mL; Thermo Fisher Scientific, after drying and purification with the SP-1

Stand Alone Solvent Purification System LC Technology Solutions Inc.) were added. To that suspension a solution of diethyl phosphite (1.35 g, 9.78 mmol, 1 eq.; Sigma-Aldrich) in THF (5 mL) was added dropwise. The mixture was subsequently heated to reflux for 5 minutes. Afterwards, the reaction mixture was cooled in an ethyl acetate-liquid nitrogen bath and C_2S (3.72 g, 47.87 mmol, 5 eq.; Sigma-Aldrich) was added to form sodium(diethoxyphosphoryl)methanedithioate (**10**). Methyl 4-(bromomethyl)benzoate (**9**; 2.46 g, 10.75 mmol, 1.1 eq.; Sigma-Aldrich) was then added dropwise to the reaction mixture and stirred for 2 h (Supplementary Figure 23; reaction (2)). Hexane was added (30 mL; Thermo Fisher Scientific), the mixture filtered and the solvent removed under vacuum conditions. The pure product (2.3 g, 6.4 mmol, 65% yield) was obtained after passing over a short silica column using first cyclohexane and then diethyl ether (Supplementary Figure 26b).

For the actual small molecule experiment (phase 3a), sorbic alcohol (reactant 1; 0.549 mg, 1 eq.) and the RAFT agent (reactant 2; 2.02 mg, 1 eq.) were dissolved in toluene (0.5 mL; Thermo Fisher Scientific, after drying and purification with the SP-1 Stand Alone Solvent Purification System LC Technology Solutions Inc.). The vial was closed and purged with a constant argon flow for 5 minutes. The reaction mixture was heated up to 353 K and maintained at that temperature. Conversion was determined using ^{1}H NMR spectroscopy (600 MHz; chloroform-d_3, Sigma-Aldrich).

Network formation (phase 3b) was subsequently performed considering a tetrafunctional and bifunctional monomeric building block related to phase 3a, respectively denoted as reactant 3 and reactant 4. Benzene-1,2,4,5-tetrayltetrakis(methylene)tetrakis((diethoxyphosphoryl)-methanedithioate) (**13**; reactant 3: tetrafunctional monomer) was synthesized as follows. In a dry 10 mL Schlenk tube with a magnetic stirrer, sodium hydride (suspension in paraffin oil; 90.0 mg, 1.74 mmol, 1 eq.; Sigma-Aldrich) was added and vigorously mixed with 3 mL heptane (Thermo Fisher Scientific). After a few minutes, the liquid phase was removed and heptane was added again. The procedure was repeated three times until a white powder of sodium hydride was obtained. The residual liquid was removed under vacuum conditions and dry THF (2 mL) was added. To that suspension a solution of diethyl phosphite (240 mg, 1.74 mmol, 1 eq.; Sigma-Aldrich) in THF (1 mL) was added dropwise. After addition, the mixture was allowed to stir for 1 h and then heated to 339 K for 2 minutes. After cooling the reaction

mixture in an acetone-liquid nitrogen bath, CS_2 (660 mg, 8.70 mmol, 5 eq.; Sigma-Aldrich) was added over a period of 1 h followed by additional stirring at ambient temperature. The reaction mixture was cooled again to 273 K and 1,2,4,5-tetrakis(bromomethyl)benzene (**12**; 195 mg, 0.42 mmol, 0.25 eq.; Sigma-Aldrich) dissolved in 1 ml dry THF was added dropwise (Supplementary Figure 25; reaction (3)). After completion, the reaction mixture was stirred for another hour, followed by solvent removal under vacuum conditions. The pure product (260 mg, 0.26 mmol, 62% yield) was obtained after flash column chromatography (100:0 to 98:2 dichloromethane:methanol); specifications: (Supplementary Figure 24e and 26c).

1,3-Bis(((2E,4E)-hexa-2,4-dien-1-yl)oxy)propan-2-ol (**15**; reactant 4: bifunctional monomer; phase 3b) was synthesized starting from sorbic alcohol. Sorbic alcohol (848 mg, 8.65 mmol, 4 eq.) and epichlorohydrin (**14**; 200 mg, 2.16 mmol, 1 eq.; Sigma-Aldrich) were placed in a 25 mL round bottom flask, equipped with a stir bar and purged with argon. To this solution, potassium hydroxide (KOH; 363 mg, 6.48 mmol, 3 eq.; Thermo Fisher Scientific) and tetrabutylammonium iodine (Bu_4NI; 160 mg, 0.43 mmol, 0.2 eq.; 98%, Sigma-Aldrich) were added (Supplementary Figure 25; reaction (4)). The reaction mixture was stirred over a period of 48 h at 338 K. After cooling to ambient temperature, the obtained slurry was dissolved in dichloromethane (DCM; 20 mL; Thermo Fisher Scientific, after drying and purification with the SP-1 Stand Alone Solvent Purification System LC Technology Solutions Inc.) and extracted with water (20 mL), followed by evaporation under reduced pressure. After flash column chromatography (93:7 to 70:30 ethyl acetate:hexane) the product was obtained as a colorless viscous oil (0.34 g, 1.34 mmol, 62% yield; (Supplementary Figure 24b, 24c and 26a).

For the actual network formation, tetrafunctional monomer (reactant 3; 26.5 mg, 0.027 mmol, 1 eq.), different equivalents of bifunctional monomer (reactant 4; 6.8 mg, 1 eq.; 10.2 mg, 1.5 eq.; 11.9 mg, 1.75 eq.; 13.6 mg, 2 eq.) and toluene (0.025 mL) were added in a 1 mL crimp vial. The vial was closed and purged with a constant argon flow for 5 minutes. Afterwards, the reaction mixture was heated up to 353 K for 24 h and analysis was performed. The networks did not dissolve in toluene and swelling tests were done in that solvent for 3 h. The percentual degree of swelling was calculated according to the ratio of 100 (w_s-w_d) and w_d, with w_s the mass of the sample after 3 h swelling and w_d the mass of sample prior

to swelling. The related experimental data are included in Fig. 4d. Images of the synthesized networks are shown in Supplementary Figure 27a (before swelling) and Supplementary Figure 27b (after addition of toluene). A clear difference in color can be seen depending on the initial ratio of the two monomers. Under stoichiometric conditions a yellow network is obtained, while going more and more to off-stoichiometric conditions (excess of tetrafunctional linker) a pink color is observed, due to the presence of unreacted dithioester moieties of the tetrafunctional linker.

In phase 3c, network synthesis was performed with Fmoc-loaded bifunctional monomer (**17**; reactant 5; 1,3-bis(((2E,4E)-hexa-2,4-dien-1-yl)oxy)propan-2-yl-4-((3-(((9H-fluoren-9-yl)methoxy)carbonyl)amino)propyl)amino)-4-oxobutanoate; short notation bifunctional linker-Fmoc), with the main results shown in Figure 5. This monomer was synthesized by modifying reactant 4 (shortly bifunctional linker-OH) in two steps (Supplementary Figure 23 reaction (5) as step 1 and reaction (6) as step 2). In step 1, sodium hydride (60% suspension in oil; 20.9 mg, 0.87 mmol, 1.1 eq.) was placed into a dry 25 mL Schlenk flask and suspended in 5 mL anhydrous THF. Under vigorous stirring reactant 3 (200.0 mg, 0.792 mmol, 1.0 eq.), solved in 3 mL THF, was added dropwise. Upon completion, the reaction mixture was allowed to stir for one additional hour at room temperature, followed by refluxing the suspension for 5 min. The reaction mixture was broad back to room temperature and succinic anhydride (87.2 mg, 0.872 mmol, 1.1 eq.; Sigma Aldrich; **1**), solved in 2 mL THF, was added dropwise. After additional stirring overnight, the solvent was removed in vacuum and the mixture evaporated onto celite. The step 1 product (shortly bifunctional linker-COOH; **16**; 0.19 g, 0.50 mmol, 63% yield) was obtained as a colorless viscous oil after flash column chromatography (90:10 to 50:50 ethyl acetate:cyclohexane with 1% formic acid; Supplementary Figure 25a and 26d).

In step 2, in a dry 10 mL Schlenk flask bifunctional linker-COOH (100 mg, 0.26 mmol, 1.0 eq.), N-Fmoc-1,3-propanediamine hydrochloride (96 mg, 0.29 mmol, 1.1 eq.; Combi-Blocks; **2**), N,N′-dicyclohexylcarbodiimide (DCC; 70 mg, 0.34 mmol, 1.3 eq.; Sigma Aldrich), and 4-dimethylaminopyridine (DMAP; 3 mg, 0.03 mmol, 0.1 eq.; Sigma Aldrich) were placed and solved in 1.5 mL dry dimethylformamide (DMF; Sigma Aldrich). After short stirring N,N-diisopropylethylamine (DIPEA; 50 µL, 0.28 mmol, 1.0 eq.; Sigma Aldrich) was added and the reaction mixture was stirred for

3 h at room temperature. Upon completion, the solvent was removed under high vacuum without applying heat. The crude reaction mixture was evaporated onto celite and purified by flash column chromatography (100:0 to 50:50 ethyl acetate:cyclohexane) to obtain bifunctional linker-Fmoc, thus reactant 5 (55 mg, 0.08 mmol, 32% yield; Supplementary Figure 25c, 25d and 26e).

The subsequent network synthesis was performed in a 1 mL crimp vial adding reactant 3 (21.0 mg, 0.022 mmol, 1.00 eq.), different equivalents of reactant 5 always obeying a 1:1 molar mixture of functionalized (bifunctional linker-Fmoc) to unfunctionalized bifunctional linker being reactant 4 (bifunctional linker-OH) (18.1 mg/7.8 mg, 2.5 eq.; 14.1 mg/6.0 mg, 2.0 eq.; 10.8 mg/4.7 mg, 1.5 eq.; 7.2 mg/3.1 mg, 1.0 eq.), and toluene (0.020 mL). The vial was closed and purged with a constant argon flow for 5 min. Afterwards, the reaction mixture was heated over 24 h at 353 K. The generated networks were removed and used for analysis. Images of the networks with the difference in color depending on the initial monomer ratio are shown in Supplementary Figure 27c.

The Fmoc (drug proxy) release was undertaken in UltraViolet-Visible (UV/Vis) cuvettes (Supplementary Figure 27e-g). The respective network (r=2.5; 1.7 mg or r=2.0; 1.8 mg or r=2.0 (20 min reaction time); 1.1 mg or r=1.5; 3.2 mg or r=1.0; 1.8 mg) was added into the cuvette and washed with DMF (three times). Afterwards, the cleavage solution (2.1 mL; 2% 1,8-diazabicyclo[5.4.0]undec-7-ene (DBU; Sigma Aldrich) in DMF) was added, the cuvette was inverted twice to ensure homogenic mixing and the UV-Vis spectrum was immediately measured (first measurement after ca. 10 s). This was repeated until the final measuring time was reached.

Part B: Main technical details of modeling framework

As conceptually illustrated in Fig. 2, we combine kinetic Monte Carlo (*k*MC) simulations with molecular dynamic (MD) simulations to follow the kinetics and 3D incorporation of each building block as a function of synthesis time *t*. In the first Supplementary methods section (Supplementary methods: general overview following flowsheets), we explain the technicalities of our framework through boxes A to N as implemented in interactive flowsheets (Supplementary Figure 1-3). A second Supplementary methods section (Supplementary Methods: detailed description of flowsheet boxes) is devoted to the

detailed description of all the steps to be undertaken in these boxes. In this subsection, which is directed toward a general reader, the most critical implementations are discussed to grasp our overall technical innovation. A specialized reader is directly referred to the Supplementary Information.

In the core matrix-based *k*MC simulations of our framework, reaction event per reaction event is sampled based on reaction probabilities represented through a cumulative probability curve. This is done starting from a sufficiently large number of initial molecules to ensure numerical convergence, as shown in Supplementary Figure 10a-d for network chemistry 1. The update of the MC reaction probabilities after an executed reaction event requires the discrete updates of the MC reaction rates, as expressed in s^{-1}. These reaction rates are in turn related to (i) the chemistry platform selected, defining the potential reaction types between the functional groups, and (ii) the reaction mixture composition, defining the (reactant) concentrations related to these reaction types.

For the calculation of intermolecular MC reaction rates, apparent rate coefficients (k_{app} values) are utilized to acknowledge that the actual observed rate can be influenced by both intrinsic chemical reactivities (k_{chem} values linked to Arrhenius parameters) and chain length and CP dependent diffusivities, implying the need of diffusion coefficients for the calculation of diffusion rate coefficients, i.e. k_{diff} values. For intramolecular MC reaction rates, the interplay of chemistry and diffusivity is also addressed by the application of fundamental distance rules to verify if the functional groups involved can react, due to a sufficient flexibility of the associated network molecule. The relevance of these diffusional limitations for both intra- and intermolecular reactions is highlighted in Supplementary Figure 12 and 13, considering network chemistry 1.

If an executed reaction event leads to the formation of a network segment or dangling chain, a composite topology matrix *T* is updated in the core *k*MC algorithm (Supplementary Figure 1). This matrix stores (i) the compositions of all individual segments and dangling chains in a main topology matrix and (ii) the associated network connectivities in 2 additional connectivity arrays embedding this matrix, therefore enabling coupling. At predetermined user-defined plotting times *T* can be employed to visualize individual (network) molecules in 2D or 3D format. For the 3D format, every row in *T* is

scanned and the associated bond length, bond angles, and dihedral angles are sampled based on cumulative probability distributions that are ideally following from MD data.

In the present work, we mainly emphasis on the organosilica network case (Fig. 2 and Fig. 3) for which Burg et al.[10] applied the MD Large-scale Atomic/Molecular Massively Parallel Simulator (LAMMPS; Supplementary Figure 3a). Post-processing allows the calculation of the equilibrium distributions of bond lengths and angles and corrections are introduced if no such MD data are available. This is for instance the case if a structural defect is encountered due to a chemical reaction omitted in typical MD simulations, as illustrated for the organosilica case in Supplementary Table 5.

Thanks to the modular nature of our modeling framework we are able to request more detailed output about molecular specifications and configurations at timestamps deemed important. Here we can rely on the general shape of e.g. the yield profile with steep variations implying a change in chemical mechanism or severity of diffusional limitations. However, this more detailed model output comes at a large computational cost. Hence, we have carefully made a compromise between guaranteeing numerical convergence and sufficient molecular detail on the one hand and an acceptable computational cost on the other hand. The computational cost amounts to a couple of hours using supercomputer platforms, at least for defect free networks. In the presence of such defects, the simulation time increases to a few days. The pre-calculated MD data has a computational cost of hours, days or weeks depending on the required potential field.

To highlight the general accuracy of our framework we showcase a successful benchmark in Supplementary Figure 11 for chain length distributions against the work of Flory, who in turn extended the work of Stockmayer.[50] Our simulated distributions for the organosilica case are consistent with the analytical equations of Flory and Stockmayer upon simplifying our framework according to their assumptions of equal functional group reactivity and the absence of intramolecular reactions.

The relevance of our modeling framework is also clear from the overview of experimental and theoretical methodologies in Supplementary Figure 5 including ca. 100 references.

Part C: Guidelines for determination of model parameters

For applying our model platform, we require a limited number of chemical and diffusion parameters. For the three network chemistries considered, these parameters are presented in Supplementary Table 1-3 according to the same outline. These model parameters are determined according to the same general guidelines, as explained in detail in the related sections in the Supplementary Discussion. There we first include an extensive discussion for the main organosilica case (Supplementary Discussion on application platform for organosilica case) and then a concise discussion in a similar fashion for the other two cases (Supplementary Discussion on epoxy-amine and Diels-Alder case). In this subsection, we highlight the common research strategy to make the reader aware that a dedicated approach has been followed and that the three sets of model parameters can be seen as reliable input values upon applying further material design, which is the core of the main text.

The common research strategy for the parameter determination is that smaller synthesis times or reaction systems with monofunctional analogues of the monomers are used to determine chemistry kinetic parameters, because such constraints allow to avoid kinetic interference of diffusional limitations. Conversely, larger synthesis times are employed to tune diffusivity parameters, considering the previously determined chemistry parameters as input. Moreover, before the actual parameter tuning, the *k*MC model is used to identify which experimental response is most sensitive to which model parameter to enable a stepwise parameter determination. Hence, our research strategy inherently minimizes parameter correlation.

For the organosilica case, the model validation is presented in Supplementary Figure 8 and 9, where the experimental NMR data are represented by symbols and the model outcome from optimized parameters (Supplementary Table 1) by full lines. A distinction is made between parameter tuning at smaller (Supplementary Figure 8) and larger times (Supplementary Figure 9). To confirm the appropriate tuning procedure simplified model descriptions are included as dashed lines. These dashed lines display clear deviations from the full lines that are the only lines consistent with the experimental data.

For the epoxy-amine curing case we utilized for the parameter tuning nIR data. The successful model validation is depicted in Supplementary Figure 20, considering the optimized parameters in Supplementary Table 2. For the Diels-Alder case again NMR data are employed. The model parameters are listed in Supplementary Table 3 and the model validation is depicted in Supplementary Figure 21.

Data availability

The data represented in Fig. 3, 4 and 5 are provided as Source data. The remaining data that support the findings of this study are included within the manuscript and its supplementary files and available from the corresponding authors upon request.

Code availability

The authors declare that the complete algorithm of the modeling framework is fully available in the Supplementary Information of the present contribution. The authors specifically refer to all implementation steps included in Supplementary Figure 1-4 and their discussion in the Supplementary Discussion complemented with information already available in the open literature.

References

1. Flory, P. J. Network topology and the theory of rubber elasticity. *Br. Polym. J.* **17,** 96-102 (1985).

2. Kryven, I. Bond percolation in coloured and multiplex networks. *Nat. Commun.* **10,** 404 (2019).

3. Zhong, M. J., Wang, R., Kawamoto, K., Olsen, B. D. & Johnson, J. A. Quantifying the impact of molecular defects on polymer network elasticity. *Science* **353,** 1264-1268 (2016).

4. Sheiko, S. S. & Dobrynin, A. V. Architectural code for rubber elasticity: from supersoft to superfirm materials. *Macromolecules* **52,** 7531-7546 (2019).

5. LLorca, J. On the quest for the strongest materials. *Science* **360,** 264-265 (2018).

6. Berman, D., Erdemir, A. & Sumant, A. V. Graphene: a new emerging lubricant. *Mater. Today* **17,** 31-42 (2014).

7. Anderson, M. W. & Klinowski, J. Direct observation of shape selectivity in zeolite ZSM-5 by magic-angle-spinning NMR. *Nature* **339,** 200-203 (1989).

8. Rimer, J. D. Rational design of zeolite catalysts. *Nat. Catal.* **1,** 488-489 (2018).

9. Boerjan, W., Ralph, J. & Baucher, M. Lignin Biosynthesis. *Annu. Rev. Plant Biol.* **54,** 519-546 (2003).

10. Burg, J. A. & Dauskardt, R. H. Elastic and thermal expansion asymmetry in dense molecular materials. *Nat. Mater.* **15,** 974-980 (2016).

11. Owens, G. J. et al. Sol–gel based materials for biomedical applications. *Prog. Mater Sci.* **77,** 1-79 (2016).

12. Loccufier, E. et al. Silica nanofibrous membranes for the separation of heterogeneous azeotropes. *Adv. Funct. Mater.* **28,** 1804138 (2018).

13. Caló, E. & Khutoryanskiy, V. V. Biomedical applications of hydrogels: a review of patents and commercial products. *Eur. Polym. J.* **65,** 252-267 (2015).

14. D'hooge, D. R., Van Steenberge, P. H. M., Reyniers, M.-F. & Marin, G. B. The strength of multi-scale modeling to unveil the complexity of radical polymerization. *Prog. Polym. Sci.* **58,** 59-89 (2016).

15. Matyjaszewski, K. Advanced materials by atom transfer radical polymerization. *Adv. Mater.* **30,** 1706441 (2018).

16. Van Steenberge, P. H. M. et al. Visualization and design of the functional group distribution during statistical copolymerization. *Nat. Commun.* **10,** 3641 (2019).

17. Di Lorenzo, F. & Seiffert, S. Nanostructural heterogeneity in polymer networks and gels. *Polym. Chem.* **6,** 5515-5528 (2015).

18. Gu, Y., Zhao, J. & Johnson, J. A. Polymer networks: from rubbers and gels to porous frameworks. *Angew. Chem. Int. Edit.* **59,** 5022-5049 (2020).

19. Sandakov, G. I., Smirnov, L. P., Sosikov, A. I., Summanen, K. T. & Volkova, N. N. NMR analysis of distribution of chain lengths between crosslinks of polymer networks. *J. Polym. Sci. Pol. Phys.* **32,** 1585-1592 (1994).

20. Mathur, A. M. & Scranton, A. B. Characterization of hydrogels using nuclear magnetic resonance spectroscopy. *Biomaterials* **17,** 547-557 (1996).

21. Tripathi, A. K., Neenan, M. L., Sundberg, D. C. & Tsavalas, J. G. Influence of n-alkyl ester groups on efficiency of crosslinking for methacrylate monomers copolymerized with EGDMA: experiments and Monte Carlo simulations of reaction kinetics and sol–gel structure. *Polymer* **96,** 130-145 (2016).

22. Kryven, I., Duivenvoorden, J., Hermans, J. & Iedema, P. D. Random graph approach to multifunctional molecular networks. *Macromol. Theor. Simul.* **25,** 449-465 (2016).

23. Fierens, S. K. et al. Model-based design to push the boundaries of sequence control. *Macromolecules* **49,** 9336-9344 (2016).

24. Buback, M., Feldermann, A., Barner-Kowollik, C. & Lacik, I. Propagation rate coefficients of acrylate-methacrylate free-radical bulk copolymerizations. *Macromolecules* **34,** 5439-5448 (2001).

25. Van Steenberge, P. H. M. et al. Linear gradient quality of ATRP copolymers. *Macromolecules* **45,** 8519-8531 (2012).

26. D'hooge, D. R. In silico tracking of individual species accelerating progress in macromolecular engineering and design. *Macromol. Rapid Commun.* **39,** 1800057 (2018).

27. Burg, J. A. et al. Hyperconnected molecular glass network architectures with exceptional elastic properties. *Nat. Commun.* **8,** 1019 (2017).

28. Schwab, F. K. & Denniston, C. Reaction and characterisation of a two-stage thermoset using molecular dynamics. *Polym. Chem.* **10,** 4413-4427 (2019).

29. Zanjani, M. B. et al. Computational investigation of the effect of network architecture on mechanical properties of dynamically cross-linked polymer materials. *Macromol. Theor. Simul.* **28,** 1900008 (2019).

30. Oliver, M. S., Dubois, G., Sherwood, M., Gage, D. M. & Dauskardt, R. H. Molecular origins of the mechanical behavior of hybrid glasses. *Adv. Funct. Mater.* **20,** 2884-2892 (2010).

31. Kilic, K. I. & Dauskardt, R. H. Design of ultrastiff organosilicate hybrid glasses. *Adv. Funct. Mater.* **29,** 1904890 (2019).

32. Talati, M., Albaret, T. & Tanguy, A. Atomistic simulations of elastic and plastic properties in amorphous silicon. *Europhys. Lett.* **86,** 66005 (2009).

33. Kawamoto, K., Zhong, M. J., Wang, R., Olsen, B. D. & Johnson, J. A. Loops versus branch functionality in model click hydrogels. *Macromolecules* **48,** 8980-8988 (2015).

34. Dušek, K. & Patterson, D. Transition in swollen polymer networks induced by intramolecular condensation. *J. Polym. Sci. Pol. Phys.* **6,** 1209-1216 (1968).

35. Panyukov, S. Loops in polymer networks. *Macromolecules* **52,** 4145-4153 (2019).

36. Mita, I. & Horie, K. Diffusion-Controlled Reactions in Polymer Systems. *J. Macromol. Sci. Polymer Rev.* **27,** 91-169 (1987).

37. Pouxviel, J. C., Boilot, J. P., Beloeil, J. C. & Lallemand, J. Y. NMR-study of the sol-gel polymerization. *J. Non-Cryst. Solids* **89,** 345-360 (1987).

38. Daelemans, L., Van Paepegem, W., D'hooge, D. R. & De Clerck, K. Excellent nanofiber adhesion for hybrid polymer materials with high toughness based on matrix interdiffusion during chemical conversion. *Adv. Funct. Mater.* **29,** 1807434 (2019).

39. Tang, J., Chen, Z., Fu, A. W. & Cheung, D. W. Capabilities of outlier detection schemes in large datasets, framework and methodologies. *Knowl. Inf. Syst.* **11,** 45-84 (2007).

40. Xia, B. et al. Preparation of silica coatings with continuously adjustable refractive indices and wettability properties via sol–gel method. *RSC Adv.* **8,** 6091-6098 (2018).

41. Yang, J., Lin, G.-S., Mou, C.-Y. & Tung, K.-L. Mesoporous silica thin membrane with tunable pore size for ultrahigh permeation and precise molecular separation. *ACS Appl. Mater. Interfaces* **12,** 7459-7465 (2020).

42. Li, J. & Mooney, D. J. Designing hydrogels for controlled drug delivery. *Nat. Rev. Mater.* **1,** 16071 (2016).

43. Oehlenschlaeger, K. K. et al. Adaptable hetero Diels-Alder networks for fast self-healing under mild conditions. *Adv. Mater.* **26,** 3561-3566 (2014).

44. Zhou, J. et al. Thermally reversible Diels-Alder-based polymerization: an experimental and theoretical assessment. *Polym. Chem.* **3,** 628-639 (2012).

45. Zhou, J., Yu, G. & Huang, F. Supramolecular chemotherapy based on host-guest molecular recognition: a novel strategy in the battle against cancer with a bright future. *Chem. Soc. Rev.* **46,** 7021-7053 (2017).

46. Harada, A., Takashima, Y. & Nakahata, M. Supramolecular polymeric materials via cyclodextrin-guest interactions. *Acc. Chem. Res.* **47,** 2128-2140 (2014).

47. Chakma, P. & Konkolewicz, D. Dynamic covalent bonds in polymeric materials. *Angew. Chem. Int. Ed.* **58,** 9682-9695 (2019).

48. Zou, W., Dong, J., Luo, Y., Zhao, Q. & Xie, T. Dynamic covalent polymer networks: from old chemistry to modern day innovations. *Adv. Mater.* **29,** 1606100 (2017).

49. St John, N. A. & George, G. A. Cure kinetics and mechanisms of a tetraglycidyl-4,4′-diaminodiphenylmethane/diaminodiphenylsulphone epoxy resin using near i.r. spectroscopy. *Polymer* **33,** 2679-2688 (1992).

50. Stockmayer, W. H. Theory of molecular size distribution and gel formation in branched-chain polymers. *J. Chem. Phys.* **11,** 45-55 (1943).

Acknowledgements

L.D.K. acknowledges the research foundation - Flanders (FWO; 1S37517N). P.H.M.V.S. and L.D. acknowledge FWO through a postdoctoral fellowship (12C4319N and 12ZR520N). D.R.D acknowledges FWO through G.0H52.16N and Vlaio (Catalisti) through the Moonshot initiative (ReSet Project). The computational resources (Stevin Supercomputer Infrastructure) and services used in this work were provided by the VSC (Flemish Supercomputer Center), funded by Ghent University, FWO and the Flemish Government - department EWI. The work at Stanford University was supported by the US Department of Energy, Office of Basic Energy Sciences, under contract no. DE-FG02-07ER46391. C.B.-K. acknowledges an Australian Research Council (ARC) Laureate Fellowship enabling his photochemical research program as well as the Queensland University of Technology (QUT) for key support. H. F. acknowledges ARC funding through a Discovery Early Career Researcher Award (DE200101096). The authors thank Eva Loccufier for discussion regarding the interpretation of the experimental data for the membrane filtration. The authors also thank Jessica Pelloth for carefully conducting initial kinetic experiments on the Diels-Alder chemistry.

Author contributions

L.D.K, P.H.M.V.S, and D.R.D. contributed to the development of matrix-based kinetic Monte Carlo (*k*MC) simulations for network synthesis. L.D.K, P.H.M.V.S., M.F.R., and D.R.D. focused on the determination of its scale dependent model parameters and contributed to the construction of the associated structure-property relationships. D.R.D. developed the overall framework of the connection of *k*MC and molecular dynamics (MD) simulations. K.I.K. and R.H.D. contributed to the part of the MD simulations and interpretation, and the construction of the structure-property relationship for the organosilica case. L.D. and K.D.C. contributed to the experimental part on the epoxy-amine curing and the construction of the related structure-property relationship. D. K., H.F. and C.B.K. contributed to the experimental part on the Diels-Alder chemistry and the construction of the related structure-property relationships. All authors have approved the manuscript and made revisions along its preparation.

Additional information

Supplementary Information accompanies this paper.

Competing interests

All authors declare no competing interests.