Extended disorder at the cell surface: the conformational landscape of the ectodomains of syndecans

Frank Gondelaud[1], Mathilde Bouakil[2†], Aurélien Le Fèvre[3†], Adriana Erica Miele[1], Fabien Chirot[3], Bertrand Duclos[1], Adam Liwo[4], Sylvie Ricard-Blum[1]*

Corresponding author: Sylvie Ricard-Blum
Email: sylvie.ricard-blum@univ-lyon1.fr

SUPPLEMENTARY DATA

Simulation procedure

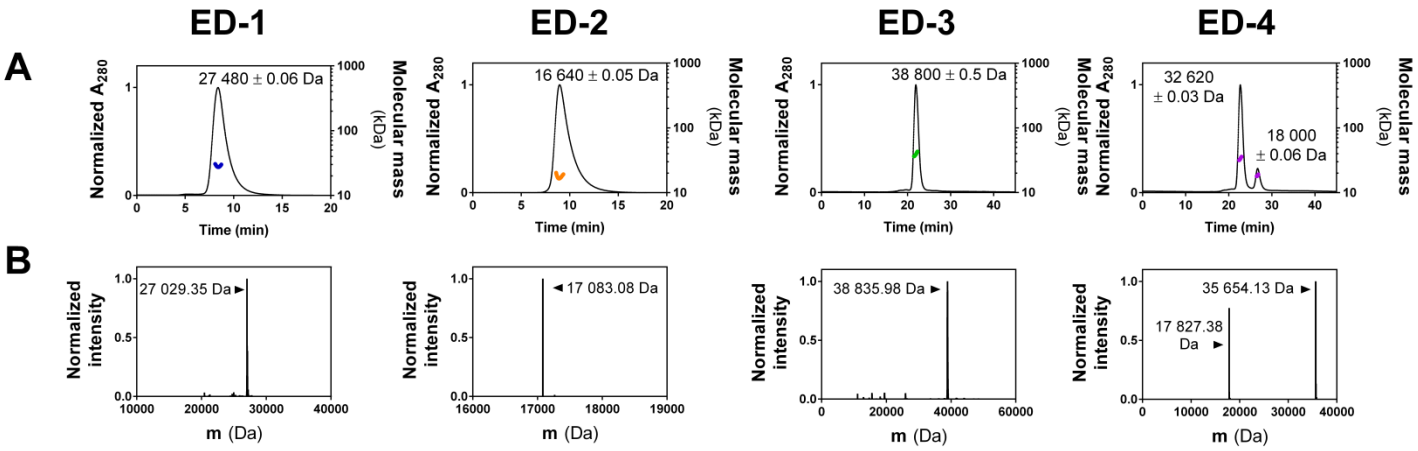
The conformational search of the C-terminal part of the syndecan ectodomain-4 dimer (a total of 80 amino-acid residues, 40 amino-acid residues per chain, residues ^{114}KRISPVEESE DVSNKVSMSSTVQGSNIFERTEVLAGCPEH^{153}) was carried out with the multiplexed replica exchange molecular dynamics method (MREMD) [1,2], which was implemented in UNRES [3] along with the molecular-dymamics protocol [4,5]. 12 replica temperatures were used, ranging from 260 K to 370 K, which were distributed to maximize the number of walks in the temperature space [6], as in our previous work [7]. 4 replicas were run at each temperature, this giving a total of 48 replicas. The simulations were started from random-generated structures of the dimer, with no pre-formed disulfide bonds. The disulfide bonds were allowed to form and break during the simulations, this process being controlled by a bimodal potential developed previously [8]. It should be noted that each chain contains one cysteine residue and, therefore, only inter-chain disulfide bond can be formed. No secondary-structure, template-based, contact or other geometry restraints were imposed.

The time step in the integration of the equations of motion was 4.89 fs. Replicas were exchanged and snapshots saved every 10,000-time steps. The temperature was controlled with the Langevin thermostat implemented with the UNRES/MD protocol [5], in which the water viscosity was scaled down by a factor of 0.02 to speed up the simulations. The simulations were carried out in a cubic periodic box [9] with the length of each side equal to 154 Å. The dimensions of the box were constant and, consequently, each replica between exchanges was run in the NVT regime (constant number of molecules, constant volume, and constant temperature). A total of 360,000,000 steps per trajectory were run until the ensemble averages were converged, which corresponds to 1.8 μs of trajectory length. However, because of averaging out the fast degrees of freedom when passing from the all-atom to a coarse-grained representation, the UNRES times scale is at least 1,000 time extended compared to the real time scale and, therefore, each trajectory corresponded to an about 1.8 ms of real time.

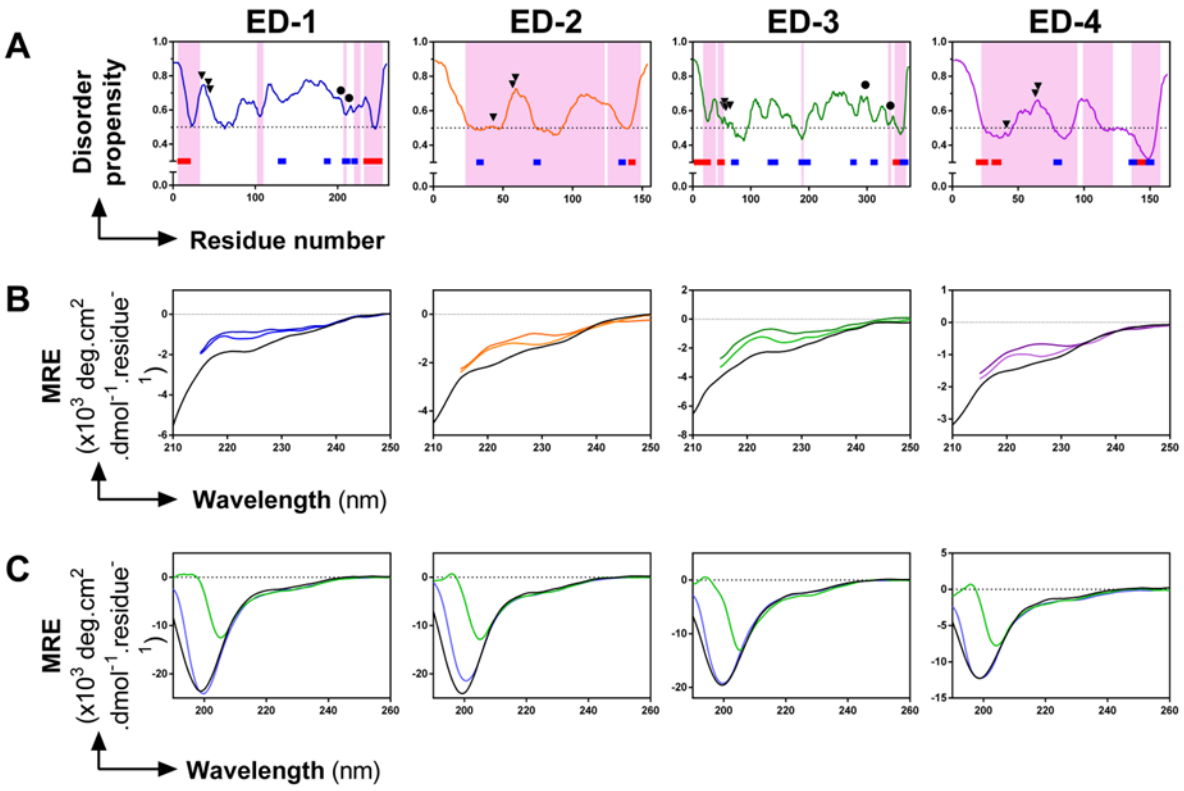
To monitor convergence, the following ensemble-averaged properties: radius of gyration, fraction of disulfide bonds, fraction of secondary structure of different type (statistical coil, β-sheet

and α-helix) were determined as functions of temperature with the use of the weighted histogram analysis method (WHAM) [10], which was implemented in UNRES in our previous work [11]. For this purpose, the simulation was divided into section, each corresponding to 40,000,000 steps per trajectory, and the plots of the above-mentioned ensemble-averaged properties from all sections were compared; the convergence was considered to be achieved when the graphs from the last sections nearly overlapped. Because WHAM is an interpolation and not an extrapolation method, we restricted the analysis to the temperature range from T=265 K (5 K above the lowest replica temperature) to T=365 K (5 K below the highest replica temperature). For each coarse-grained structure contributing to the ensemble, the radius of gyration was calculated from coarse-grained structures, based on the positions of the C^{α} atoms. The content of secondary structure of a structure was also calculated from coarse-grained conformations, based on the algorithm of the detection of secondary structure from coarse-grained structures developed in our earlier work [12].

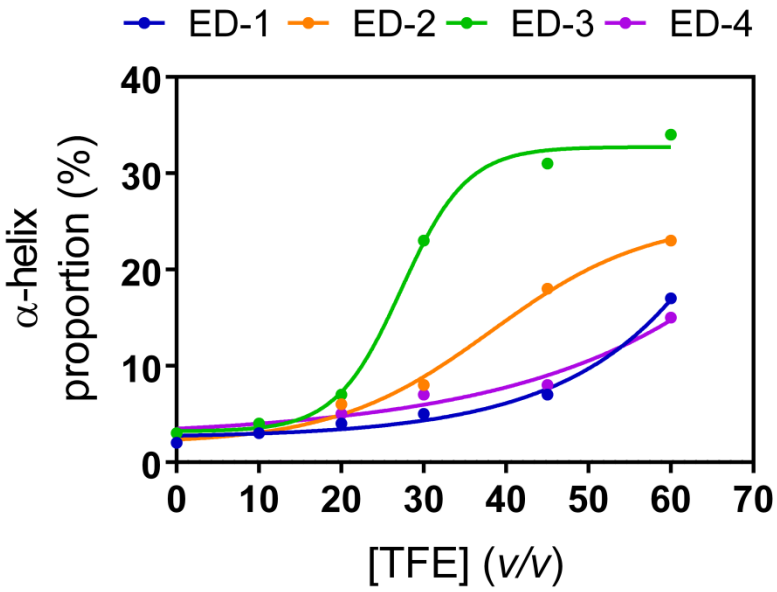
To determine the representative conformations, we followed the procedure developed in an earlier work [13]. First, a cluster analysis was carried out, taking the conformations which, according to their probabilities at the temperature of choice (300 K in this work) as calculated by WHAM[11] constituted 99% of the ensemble. The cluster analysis was performed with Ward's minimum variance method [14]. We set the number of clusters to 5. The population of each cluster was determined as a sum of probabilities of all constituent conformations [11]. Subsequently, the representative conformation of each cluster was selected as the conformation closest the cluster weighted average (with the weights being the probabilities calculated by WHAM). The information of the presence of absence of a disulfide bond was also included. In the next step, the conformations were converted to all-atom representation with the use of the PULCHRA [15] and SCWRL [16] algorithms and refined by doing a short minimization followed by a short MD simulation with the AMBER14 package that uses the ff99SB force field [17].



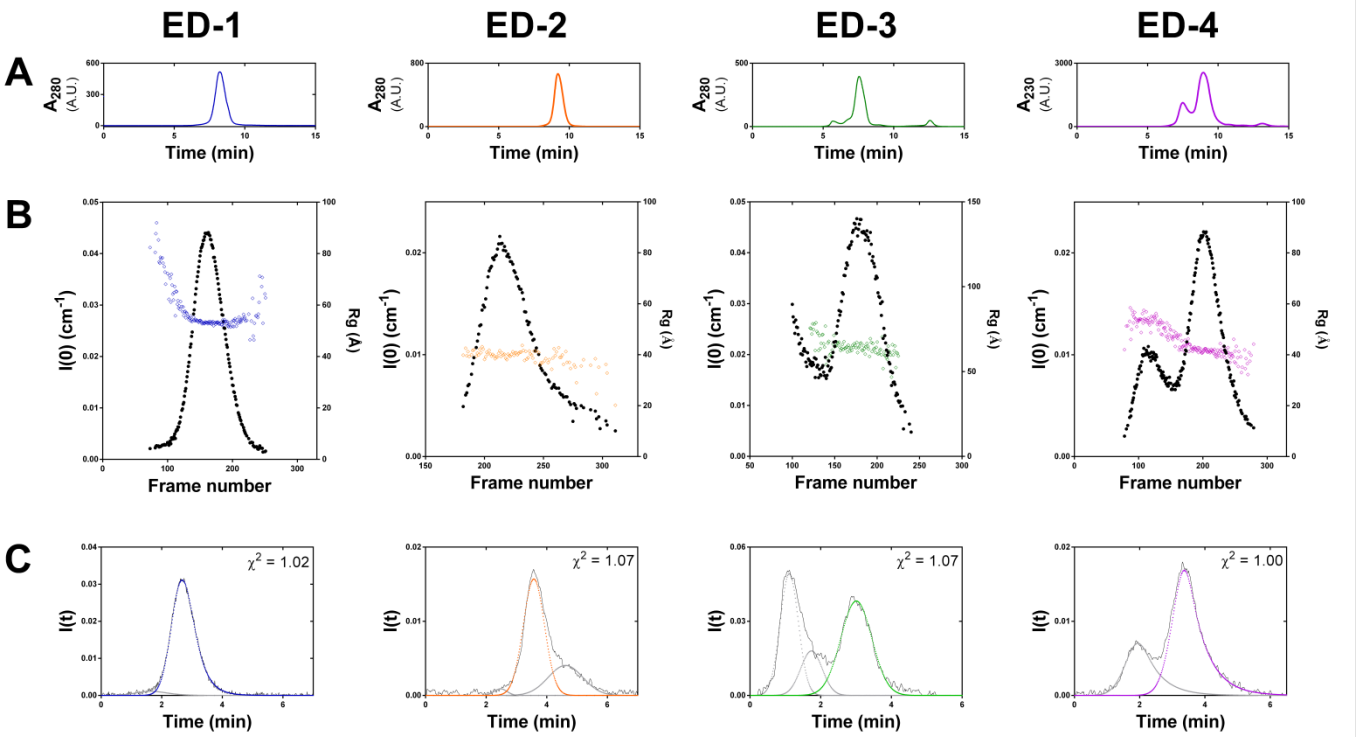
Supplementary Fig. S1. **Molecular mass determination of the syndecan ectodomains by SEC-MALLS and ESI-MS. A)** Analysis of the four ectodomains of syndecans (EDs) by SEC-MALLS. UV spectra (black lines) and molecular masse estimation (colored curves) are represented. **B)** ESI-MS spectra of the four syndecan ectodomains.



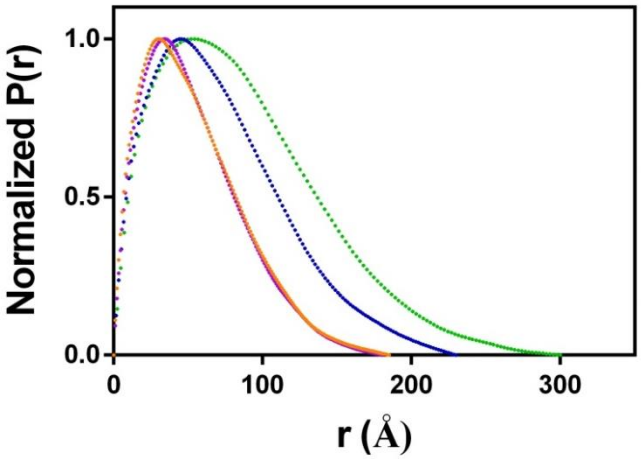
Supplementary Fig. S2. Intrinsic disorder, molecular recognition features (MoRFs) and secondary structures of the four ectodomains of human syndecans. A) The disorder propensity of the tagged recombinant ectodomains was predicted with the metaPrDOS predictor [18] (false positive rate: 5%, propensity cut-off: 0.5). The disordered residues are displayed as a curve, and the disorder cut-off (0.5) is visualized by a black dotted line. Molecular recognition features, represented as pink areas, were predicted with MoRFchibi SYSTEM based on the propensity of amino acid residues to undergo a disorder-to-order transition upon binding to a globular protein [19]. The secondary structure was predicted using PSIPRED 4.0 with default parameters [20]. Red and blue rectangles: predicted α-helix and β-strands respectively. Heparan sulfate and chondroitin sulfate chains attachment sites on the ectodomains are represented as black triangles and circles respectively. The amino acid residues are numbered using the sequence of the recombinant syndecan ectodomains, including the 6 Histidine tag at the N-terminus, and the FLAG tag at the C-terminus. **B)** Circular dichroism (CD) spectra of the syndecan ectodomains in absence (black line) and in presence of 3 and 5 M (the darkest color) guanidinium chloride. The spectra were recorded between 260 and 215 nm in presence of guanidinium chloride because of the high absorbance of chloride ions. **C)** CD spectra of the syndecan ectodomains in absence (black spectra) or in presence of 0.5 M trimethylamine N-oxide (green spectra) or of 100 µM 4-phenylbutyric acid (blue spectra) in 10 mM potassium phosphate pH 7.4. The deconvolution of CD spectra using Dichroweb [21] was performed as described in the Methods section.



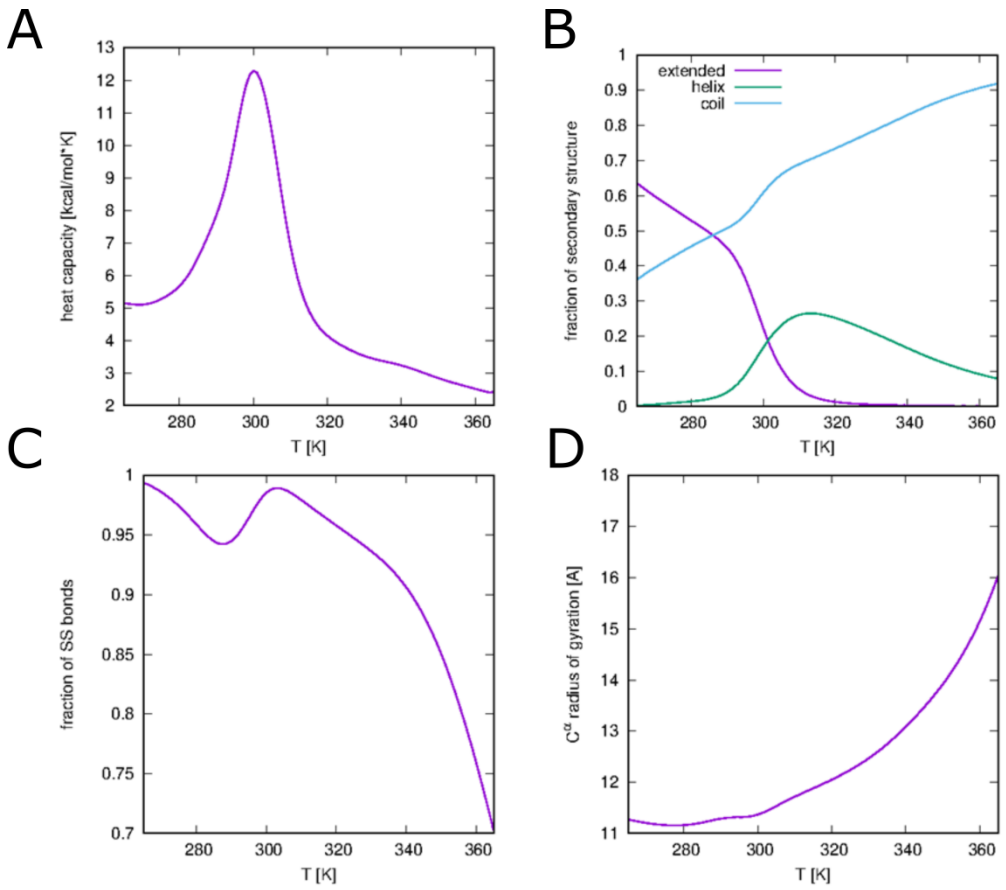
Supplementary Fig. S3. Formation of α-helix in the ectodomains of syndecans (ED) in presence of increasing concentrations of trifluoroethanol (TFE). The respective proportions of α-helix were extracted from the deconvolution of CD spectra as detailed in the online Methods section. A sigmoidal fitting was performed for each dataset (goodness-of-fit (R^2) ≥ 0.99 for ED-1, ED-2 and ED-3 and 0.97 for ED-4).



Supplementary Fig. S4. SEC-SAXS data processing. A) SEC elution profiles of the four ectodomains of syndecans (EDs). **B)** I(0)/Rg as a function of the recorded frames. Black dots: I(0) intensities, colored diamonds: estimation of the radius of gyration. **C)** Deconvolution of the scattering intensities using the US-SOMO HPLC SAXS module [22]. The colored Gaussian curves correspond to the monomeric form of the ectodomains. Molecular species of ED-4 eluted at around 2 min correspond to the dimeric form of this ectodomain.



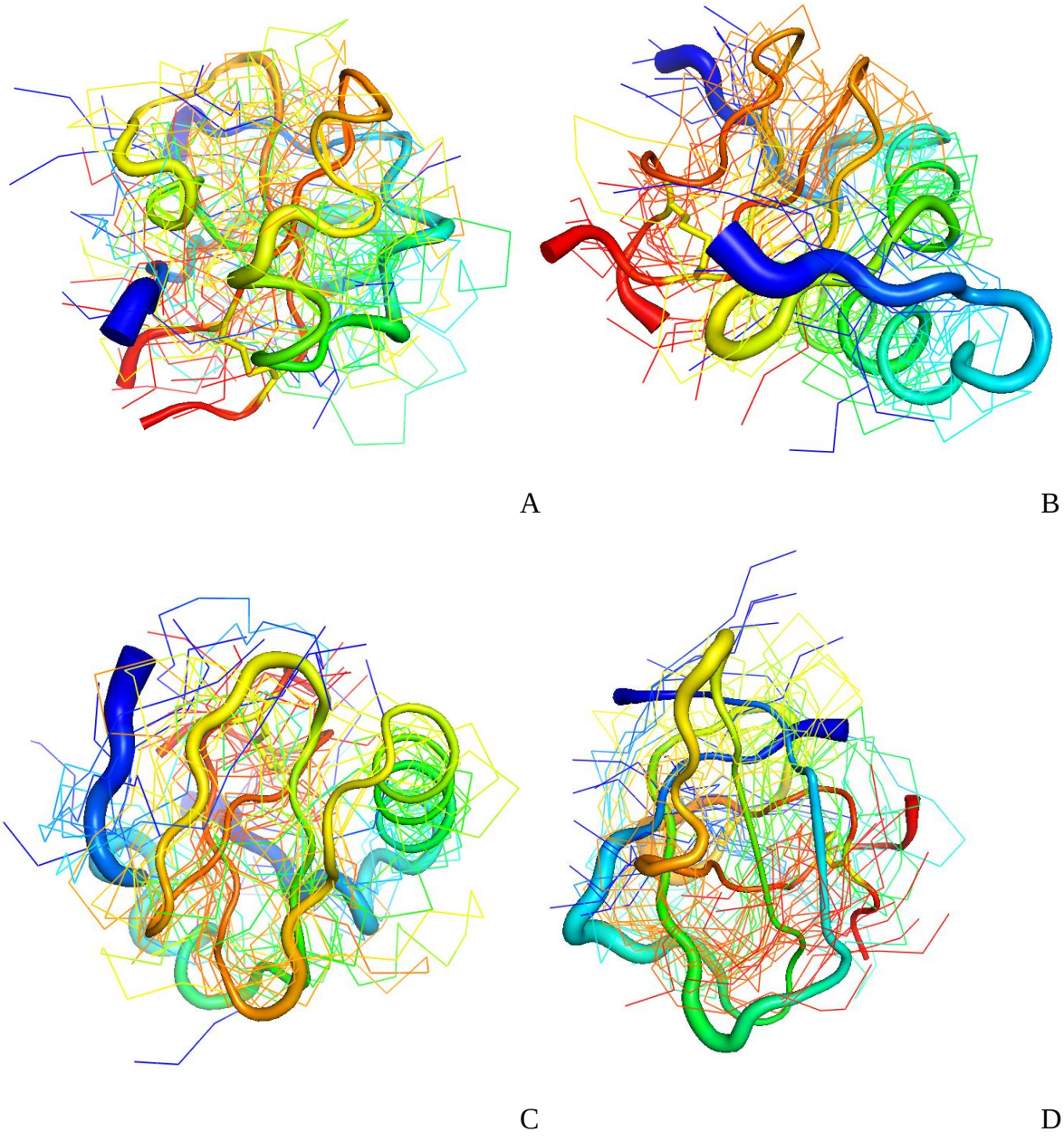
Supplementary Fig. S5. Pair-distance distribution function P(r) calculated from SAXS intensities using GNOM implemented in the ATSAS package. Curves were normalized and adjusted until obtaining a smooth decay to zero. The quality of P(r) was assessed with the CorMap test. ED-1: blue, ED2: orange, ED-3: green, ED-4: purple. The curves obtained for ED-2 and ED-4 are similar, and appear to be superimposed.



Supplementary Fig. S6. Plots of temperature dependence of the heat capacity. (**A**), ensemble-averaged fractions of secondary structure of different type (**B**), ensemble-averaged fraction of disulfide bonds (**C**), and ensemble-averaged radius of gyration calculated over C^{α} atoms (**D**) of the C-terminal part of syndecan-4 ectodomain, obtained from the last (320,000,000-360,000,000) sections of the MREMD trajectories with WHAM. The graphs were made with gnuplot (Williams T, Kelley C et al., Gnuplot 4.6: an interactive plotting program, 2013, http://gnuplot.sourceforge.net).

It can be seen from **Supplementary Fig. S6A** that a clear phase transition occurs at T=300 K. The position of this peak nearly coincides with the temperature of T=301 K, at which the fraction of β and helical structures become equal and the fraction of coil structure (**Supplementary Fig. S6B**). At this point, the fractions of extended, helix, and coil structures are both 19 % and that of the coil structure is 62 %. At lower temperatures, the β-sheet structure dominates. The fraction of α-helical structure increases until T=314 K then decreases with temperature; however, even at T=365 K there is about 10 % of helical structure. The fraction of disulfide bonds (**Supplementary Fig. S6C**) is over 90 % until T=341 K, then sharply drops to about 70 % at T=365 K, which is short of water-boiling temperature. It can, therefore, be concluded that the system consists of disulfide-bonded dimers even at denaturing conditions when the secondary structure is already largely lost. There

are some minor fluctuations of the radius of gyrations until T=300 K but, because they amount to a few percent only, we refrain from discussing them and drawing possible conclusions here. From **Supplementary Fig. S6D** it can be seen that the mean dimension of the system, quantified as the radius of gyration computed from C^{α}-positions, remains nearly constant (about 11.3 Å) until T=300 K and then starts to increase to grow to 16.0 Å (by about 50%) at T=365 K. However, no inflection point is observed in the analyzed range of temperatures (which extends to that of nearly boiling water).



Supplementary Fig. S7. The representative conformations of the four dominant families 1-4 at T=300 K (A-D), with the representative structure shown in the tube representation. The tube radius corresponds to the estimated relative coordinate uncertainty of the respective region, and 10 leading conformations of a family are shown in line representation. The orientations of the representative conformations are as in Supplementary Fig. S6. Each chain is colored from blue to red from the N- to the C-terminus. The structures were drawn with PyMOL (The PyMOL Molecular Graphics System, Version 1.8.2.1, Schrödinger, LLC).

Recombinant protein	UniProtKB Accession number	Residues (UniProtKB)	Number of amino acid residues including tags
ED-1	P18827	23 - 254	260
ED-2	P34741	19 - 144	154
ED-3	O75056	47 - 387	373
ED-4	P31431-2	19 - 153	163
ED-4 C150A	-	19 - 153	163

Supplementary Table S1. Characteristics of the recombinant ectodomains (EDs) of human syndecans expressed with a 6xHis tag at the N-terminus, and a FLAG tag at the C-terminus.

Syndecan ectodomains	Molecular weight of tagged recombinant EDs (Da)				
	Theoretical	**SDS-PAGE**	**SEC**	**ESI-MS**	**SEC-MALLS**
ED-1	27 030	42 300	172 000	27 029.4	27 480 ± 60
ED-2	17 083	27 700	72 000	17 083.1	16 640 ± 50
ED-3	38 831	89 000	389 000	38 836.0	38 800 ± 500
ED-4	17 828	27 500 63 300	78 000 195 000	17 827.4 35 654.1	18 000 ± 60 32 620 ± 30
ED-4 C150A	17 790	27 300	91 000	17 795.6	17 592 ± 50

Supplementary Table S2. Apparent molecular weights of the recombinant tagged ectodomains of human syndecans calculated by SDS-PAGE and size-exclusion chromatography using globular proteins as standards, and molecular masses determined experimentally by ESI-MS and SEC-MALLS. The proteins were expressed with a N-terminal 6xHis tag and a C-terminal FLAG tag.

		ED-1	ED-2	ED-3	ED-4 (monomer)
Average conformations					
$SASA_{av}$ ($Å^2$)		29 653.0	16 074.2	28 624.4	17 780.2
Extended conformations					
$SASA_{ext}$ ($Å^2$)		31 103.9	18 268.6	42 071.3	18 843.5
Compact conformations					
Average charge	EDs	11.69	7.81	15.24	8.35
	Globular protein	10.54	8.70	13.89	8.65
	Difference	+ 1.15	- 0.89	+ 1.35	- 0.30
$SASA_{cpt}$ ($Å^2$)	EDs	12 912.8	7 090.0	19 178.1	7 832.4
	Globular protein	10 997	8 017	15 586	8 321
Number of amino acid residues	EDs	260	154	373	163
	Globular proteins	260 (Carbonic anhydrase 1)	153 (Myoglobin)	386 (Ovalbumin)	163 (β-lacto-globulin A)

Supplementary Table S3. Solvent-accessible surface area (SASA) of the syndecan ectodomains (EDs) calculated from native ESI-MS experiments. The average SASA ($SASA_{av}$) was calculated using the entire charge state distribution whereas the extended ($SASA_{ext}$) and compact ($SASA_{cpt}$) SASAs were calculated for the most and less charged envelopes, respectively. ED-4 SASAs refer to its monomeric form.

References

[1] U.H.E. Hansmann, Y. Okamoto, Comparative study of multicanonical and simulated annealing algorithms in the protein folding problem, Physica A: Statistical Mechanics and Its Applications. 212 (1994) 415–437. https://doi.org/10.1016/0378-4371(94)90342-5.

[2] Y.M. Rhee, V.S. Pande, Multiplexed-Replica Exchange Molecular Dynamics Method for Protein Folding Simulation, Biophys J. 84 (2003) 775–786.

[3] C. Czaplewski, S. Kalinowski, A. Liwo, H.A. Scheraga, Application of Multiplexed Replica Exchange Molecular Dynamics to the UNRES Force Field: Tests with alpha and alpha+beta Proteins, J Chem Theory Comput. 5 (2009) 627–640. https://doi.org/10.1021/ct800397z.

[4] M. Khalili, A. Liwo, F. Rakowski, P. Grochowski, H.A. Scheraga, Molecular dynamics with the united-residue model of polypeptide chains. I. Lagrange equations of motion and tests of numerical stability in the microcanonical mode, J Phys Chem B. 109 (2005) 13785–13797. https://doi.org/10.1021/jp058008o.

[5] M. Khalili, A. Liwo, A. Jagielska, H.A. Scheraga, Molecular dynamics with the united-residue model of polypeptide chains. II. Langevin and Berendsen-bath dynamics and tests on model alpha-helical systems, J Phys Chem B. 109 (2005) 13798–13810. https://doi.org/10.1021/jp058007w.

[6] S. Trebst, M. Troyer, U.H.E. Hansmann, Optimized parallel tempering simulations of proteins, J Chem Phys. 124 (2006) 174903. https://doi.org/10.1063/1.2186639.

[7] A. Liwo, A.K. Sieradzan, A.G. Lipska, C. Czaplewski, I. Joung, W. Żmudzińska, A. Hałabis, S. Ołdziej, A general method for the derivation of the functional forms of the effective energy terms in coarse-grained energy functions of polymers. III. Determination of scale-consistent backbone-local and correlation potentials in the UNRES force field and force-field calibration and validation, J Chem Phys. 150 (2019) 155104. https://doi.org/10.1063/1.5093015.

[8] M. Chinchio, C. Czaplewski, A. Liwo, S. Ołdziej, H.A. Scheraga, Dynamic Formation and Breaking of Disulfide Bonds in Molecular Dynamics Simulations with the UNRES Force Field, J Chem Theory Comput. 3 (2007) 1236–1248. https://doi.org/10.1021/ct7000842.

[9] A.K. Sieradzan, Introduction of periodic boundary conditions into UNRES force field, J Comput Chem. 36 (2015) 940–946. https://doi.org/10.1002/jcc.23864.

[10] S. Kumar, J.M. Rosenberg, D. Bouzida, R.H. Swendsen, P.A. Kollman, THE weighted histogram analysis method for free-energy calculations on biomolecules. I. The method, Journal of Computational Chemistry. 13 (1992) 1011–1021. https://doi.org/10.1002/jcc.540130812.

[11] A. Liwo, M. Khalili, C. Czaplewski, S. Kalinowski, S. Ołdziej, K. Wachucik, H.A. Scheraga, Modification and optimization of the united-residue (UNRES) potential energy function for canonical simulations. I. Temperature dependence of the effective energy function and tests of the optimization method with single training proteins, J Phys Chem B. 111 (2007) 260–285. https://doi.org/10.1021/jp065380a.

[12] S. Ołdziej, J. Łągiewka, A. Liwo, C. Czaplewski, M. Chinchio, M. Nanias, H.A. Scheraga, Optimization of the UNRES Force Field by Hierarchical Design of the Potential-Energy Landscape. 3. Use of Many Proteins in Optimization, J. Phys. Chem. B. 108 (2004) 16950–16959. https://doi.org/10.1021/jp040329x.

[13] P. Krupa, M.A. Mozolewska, M. Wiśniewska, Y. Yin, Y. He, A.K. Sieradzan, R. Ganzynkowicz, A.G. Lipska, A. Karczyńska, M. Ślusarz, R. Ślusarz, A. Giełdoń, C.

Czaplewski, D. Jagieła, B. Zaborowski, H.A. Scheraga, A. Liwo, Performance of protein-structure predictions with the physics-based UNRES force field in CASP11, Bioinformatics. 32 (2016) 3270–3278. https://doi.org/10.1093/bioinformatics/btw404.
[14] F. Murtagh, A. Heck, Multivariate Data Analysis, Springer Netherlands, 1987. https://doi.org/10.1007/978-94-009-3789-5.
[15] P. Rotkiewicz, J. Skolnick, Fast procedure for reconstruction of full-atom protein models from reduced representations, J Comput Chem. 29 (2008) 1460–1465. https://doi.org/10.1002/jcc.20906.
[16] Q. Wang, A.A. Canutescu, R.L. Dunbrack, SCWRL and MolIDE: computer programs for side-chain conformation prediction and homology modeling, Nat Protoc. 3 (2008) 1832–1847. https://doi.org/10.1038/nprot.2008.184.
[17] J.A. Maier, C. Martinez, K. Kasavajhala, L. Wickstrom, K.E. Hauser, C. Simmerling, ff14SB: Improving the Accuracy of Protein Side Chain and Backbone Parameters from ff99SB, J Chem Theory Comput. 11 (2015) 3696–3713. https://doi.org/10.1021/acs.jctc.5b00255.
[18] T. Ishida, K. Kinoshita, Prediction of disordered regions in proteins based on the meta approach, Bioinformatics. 24 (2008) 1344–1348. https://doi.org/10.1093/bioinformatics/btn195.
[19] N. Malhis, M. Jacobson, J. Gsponer, MoRFchibi SYSTEM: software tools for the identification of MoRFs in protein sequences, Nucleic Acids Res. 44 (2016) W488-493. https://doi.org/10.1093/nar/gkw409.
[20] D.W.A. Buchan, D.T. Jones, The PSIPRED Protein Analysis Workbench: 20 years on, Nucleic Acids Res. 47 (2019) W402–W407. https://doi.org/10.1093/nar/gkz297.
[21] L. Whitmore, B.A. Wallace, Protein secondary structure analyses from circular dichroism spectroscopy: methods and reference databases, Biopolymers. 89 (2008) 392–400. https://doi.org/10.1002/bip.20853.
[22] E. Brookes, P. Vachette, M. Rocco, J. Pérez, US-SOMO HPLC-SAXS module: dealing with capillary fouling and extraction of pure component patterns from poorly resolved SEC-SAXS data, J Appl Crystallogr. 49 (2016) 1827–1841. https://doi.org/10.1107/S1600576716011201.