Table of Contents

List of Figures

List of Tables

Author Contributions

Experiments were designed by JBC and NJO. Cloning, expression and purification (see Biochemical Methods) of PksJ ACP4 and mutants was performed by MP and MJ. Mass Spectrometry experiments were performed by LO and JBC. Python code development and Molecular Dynamics simulations were performed by JBC. The manuscript was written by JBC, revised by JBC and NJO with comments from LO, MP and MJ.

Experimental Details

The following are the complete experimental details for: the cloning, expression and purification; mass spectrometry and collision induced unfolding; and molecular dynamics of PksJ ACP4 and its alanine mutants.

Biochemical Methods
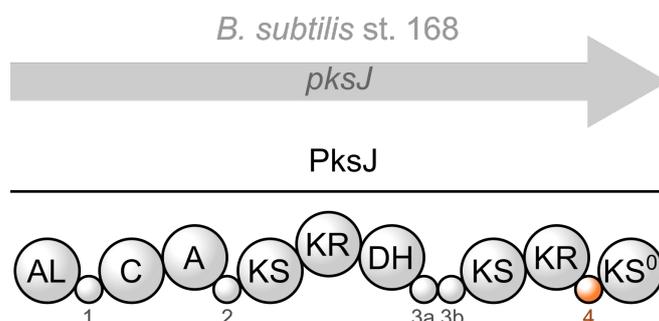
Figure S1 | Schematic of PksJ[1] protein showing the defined domains of the PKS. Enzymatic domains (shown as circles) are given abbreviations: AL, acyl ligase; C, condensation domain; A, adenylation domain; KS, ketosynthase; KR, ketoreductase; DH, dehydratase. The ACPs are shown as small circles with the ACP number below, they are numbered sequentially by appearance in the PKS, tandem ACPs are given alphabetical suffixes. Under traditional nomenclature: ACP2 is called PCP1, ACPs 3 and 4 are 2 and 3, respectively. PksJ ACP4 (used in this study) is highlighted in orange.

Cloning and Mutagenesis

The amplification of PksJ ACP4 domain from *B. subtilis* st. 168 gDNA[1] was performed using Q5 DNA polymerase (NEB) using the following primers (restriction enzyme sites are highlighted in bold):

PksJ_ACP4_For   (5'-ATA**GGATCC**GCCGATGAAGTTTCCAAATG-3')
PksJ_ACP4_Rev   (5'-ATA**CTCGAT**CAAGGTGCTGATGCATCCGAT-3')

The PCR product was separated on a 1% agarose gel and the band was excised and purified with a GeneJet gel extraction kit (Thermo Scientific). The purified inserts were digested using BamHI and XhoI, and subsequently ligated with pET28a(+) using T4 ligase (NEB) which had been pre-digested with the corresponding restriction enzymes. The ligation mixture was used to transform E. coli TOP10 cells (Invitrogen), which were plated on LB agar containing kanamycin (50 μg/mL). Colonies were picked and grown overnight in LB media containing (50 μg/mL). Plasmids were isolated from cultures using a mini-prep kit (Thermo), and the insert was sequenced to verify integrity.

All PksJ ACP4 alanine scanning mutants were constructed using the Q5 site-directed mutagenesis kit (NEB), using primers detailed in Table S1 introducing a (YYY→GTC, X→Ala) mutation in each case. PCR products were processed according to the manufactures protocol, and resulting plasmids sequenced to verify correct mutation.

The expected amino-acid sequence of the wild-type PksJ ACP4 is:

```
        -29         -19         -9          1           11
     mgss hhhhhhssgl vprgshmasm tggqqmgrgs ADEVSKCDGL LSETQSWLID
21          31          41          51          61          71
LFTEELRIDR EDFEIDGLFQ DYGVDSIILA QVLQRINRKL EAALDPSILY EYPTIQRFAD
81          91          101
WLIGSYSERL SALFGGRISD ASAP
```

Where the residues are numbered from the beginning of the inserted gene—this numbering is used to describe the mutants herein. Mutants used in this study are highlighted.

Table S1 | Primers and associated annealing temperatures (T$_a$) for construction of the 34 PksJ ACP4 mutants.

Mutant	Forward Primer (5'-3')	Reverse Primer (5'-3')	T$_a$ (°C)
E13A	ATTATTATCTGCTACACAGTCCTGG	CCATCACATTTGGAAACTTC	56
Q15A	ATCTGAAACAGCTTCCTGGCTTATTGATCTGTTTAC	AATAATCCATCACATTTGGAAAC	58
D20A	CTGGCTTATTGCTCTGTTTACCG	GACTGTGTTTCAGATAATAATC	57
E24A	TCTGTTTACCGCTGAGCTGAGAATAG	TCAATAAGCCAGGACTGTG	59
E25A	GTTTACCGAAGCTCTGAGAATAGATC	AGATCAATAAGCCAGGAC	56
L26A	TACCGAAGAGGCTAGAATAGATCGTGAAGAC	AACAGATCAATAAGCCAG	57
R27A	CGAAGAGCTGGCTATAGATCGTGAAG	GTAAACAGATCAATAAGCC	56
I28A	AGAGCTGAGAGCTGATCGTGAAGAC	TCGGTAAACAGATCAATAAG	56
D29A	GCTGAGAATAGCTCGTGAAGACTTC	TCTTCGGTAAACAGATCAATAAG	59
R30A	GAGAATAGATGCTGAAGACTTCGAGATTGACG	AGCTCTTCGGTAAACAGATC	59
E31A	AATAGATCGTGCTGACTTCGAGATTG	CTCAGCTCTTCGGTAAAC	57
D32A	AGATCGTGAAGCTTTCGAGATTGACG	ATTCTCAGCTCTTCGGTAAAC	59
F33A	TCGTGAAGACGCTGAGATTGACGGG	TCTATTCTCAGCTCTTCG	58
E34A	TGAAGACTTCGCTATTGACGGGTTG	CGATCTATTCTCAGCTCTTC	60
D36A	CTTCGAGATTGCTGGGTTGTTTCAG	TCTTCACGATCTATTCTCAG	57
Q40A	CGGGTTGTTTGCTGATTATGGCG	TCAATCTCGAAGTCTTCAC	57
D41A	GTTGTTTCAGGCTTATGGCGTGG	CCGTCAATCTCGAAGTCTTC	63
Y42A	GTTTCAGGATGCTGGCGTGGATTC	AACCCGTCAATCTCGAAG	59
Q51A	CATTTTGGCAGCTGTGCTCCAGCGTATAAAC	ATCGAATCCACGCCATAATC	62
Q54A	GGCACAGGTGGCTCAGCGTATAAAC	AAAATGATCGAATCCACG	57
R55A	ACAGGTGCTCGCTCGTATAAACCG	GCCAAAATGATCGAATCC	59
R58A	GGTGCTCCAGGCTATAAACCGCAAATTAG	TGTGCCAAAATGATCGAATC	64
K59A	GCGTATAAACGCTAAATTAGAGGCAGCGCTCG	TGGAGCACCTGTGCCAAA	69
E61A	TATAAACCGCGCTTTAGAGGCAGCGCTCGATCCATC	CGCTGGAGCACCTGTGCC	64
D65A	CCGCAAATTAGCTGCAGCGCTCG	TTTATACGCTGGAGCACC	67
P66A	GGCAGCGCTCGCTCCATCGATTC	TCTAATTTGCGGTTTATACGCTGG	63
S67A	AGCGCTCGATGCTTCGATTCTATATG	GCCTCTAATTTGCGGTTTATAC	62
Y70A	GCTCGATCCAGCTATTCTATATGAATACC	GCTGCCTCTAATTTGCGG	67
E71A	ATCGATTCTAGCTGAATACCCGACAATTCAAAGGTTCGC	GGATCGAGCGCTGCCTCT	66
Y72A	GATTCTATATGCTTACCCGACAATTCAAAGGTTCGC	GATGGATCGAGCGCTGCC	62
Q76A	TCTATATGAAGCTCCGACAATTCAAAGGTTCGC	ATCGATGGATCGAGCGCT	58
R77A	GACAATTCAAGCTTTCGCAGATTGGCTGATC	GGGTATTCATATAGAATCGATG	58
D80A	AAGGTTCGCAGCTTGGCTGATCG	TGAATTGTCGGGTATTCATATAGAATC	64
W81A	GTTCGCAGATGCTCTGATCGGTTC	CTTTGAATTGTCGGGTATTC	58

Protein Overproduction and Purification

A single colony of *E. coli* BL21(DE3) that had been transformed with the appropriate expression vector was picked and used to inoculate LB medium (5 or 10 mL) containing kanamycin (50 µg/mL). The resulting culture was incubated overnight at 37°C and 180 rpm then used to inoculate LB medium (0.5 or 1 L) containing kanamycin (50 µg/mL). The resulting culture was incubated at 37°C and 180 rpm until the optical density of the culture at 595 nm reached 0.6, then IPTG (1 mM) was added and growth was continued overnight at 15°C and 180 rpm. The cells were harvested by centrifugation (4,000 g, 15 min, 4°C) and re-suspended in buffer (20 mM Tris-HCl, 100 mM NaCl, 20 mM Imidazole, pH 7.4) at 10 mL/L of growth medium then lysed using a Constant Systems cell disruptor.

The lysate was centrifuged (37,000 g, 30 min, 4°C) and the resulting supernatant was loaded onto a HiTrap FF Chelating Column (GE Healthcare), which had been pre-loaded with 100 mM NiSO$_4$ and equilibrated in re-suspension buffer (20 mM Tris-HCl, 100 mM NaCl, 20 mM Imidazole, pH 7.4). Proteins were eluted in a stepwise manner using re-suspension buffer containing increasing concentrations of imidazole — 50 mM (5 mL), 100 mM (3 mL), 200 mM (3 mL) and 300 mM (3 mL). The presence of the protein of interest in fractions was confirmed by SDS-PAGE; an additional gel filtration step (Superdex 75/200, GE Healthcare) was used to further purify proteins where necessary. Fractions containing the protein of interest were pooled and concentrated to 250-400 µM using a Viva-Spin MWCO 10000 centrifugal concentrator (Sartorius). Samples were snap-frozen in liquid N$_2$ and stored at -80°C.

Mass Spectrometry

All mass spectrometry experiments were performed on a Synapt G1 HDMS instrument (Waters, Wilmslow, UK) in positive ion mode. Dry N$_2$ gas was used for desolvation and ion mobility gases while dry argon was used as the collision gas in the trap and transfer collision cells. See Table S2 below for details of instrumental conditions. All data files were acquired over the 400–4000 *m/z* range.

Sample Preparation

Aliquots of purified PksJ ACP4 wild-type and mutants (10 µL at 50 µM) were thawed on ice before exchange into ammonium acetate (50 mM) via Zeba Desalting cartridges (75 µL, 7k MWCO) (ThermoFisher, Hemel Hempstead, UK) following the manufacturer's instructions. Briefly, the cartridges were equilibrated with 4 × 50 µL ammonium acetate by centrifugation (1,000 g, 1 min, 8°C) before carefully loading the 10 µL protein sample onto the resin and collecting by centrifugation (1,000 g, 2 min, 8°C). The exchanged sample was then diluted to 10 µM with ammonium acetate.

Collision Induced Unfolding of ACP4

The exchanged and diluted ACP4 was directly infused via a Hamilton syringe (100 µL) at 5 µL/min into a standard Waters z-spray source. Optimised instrument conditions are listed in Table S2, these conditions were tuned initially with native myoglobin and cytochrome c before fine tuning with wild-type ACP4. Initial full scan spectra were recorded at low collision voltage. Quadrupole isolation with a narrow window was applied to the 8+ and 7+ charge states to ensure adducts did not interfere with analysis (such as via charge stripping). The ions were then activated in the Trap region of the instrument with collision voltages of 5–20 V and 15–30 for the 8+ and 7+, respectively; increments

were initially 1 V with additional measurements at 0.5 V increments recorded at 10–16 V and 18–23 V for the 8+ and 7+, giving 23 and 21 total measurements respectively.

Table S2 | Instrument conditions for the SYNAPT HDMS that may affect ion mobility and protein stability.

Pre-IMS Energies		T-Wave Voltages		Gases		Pressures (mbar)	
Capillary	2.5 kV	Trap Wave	300 m/s, 0.2 V	Cone	30 L/hr	Backing	3.78
Sampling Cone	10 V	IMS Wave	280 m/s, 7 V	Desolvation	50 L/hr, 50°C	Source	2.26×10^{-3}
Extraction Cone	5 V	Transfer Wave	200 m/s, 7 V	Trap	4 mL/min	Trap	2.95×10^{-2}
Trap Collision	5–30 V	Trap Bias	10 V	IMS	25 mL/min	IMS	4.28×10^{-1}
Source Temperature	50°C	Trap Trap Height	15 V				

Analysis of CIU Data

Arrival time distributions were extracted from MassLynx 4.1 data files for a 3 *m/z* window around the quad isolated peak. These were compiled into CSV file compatible with CIUSuite2.[2,3] CIU fingerprint plots were generated using CIUSuite2 and $CIU50_{FD}$ calculated using the feature detection tool (CIU50 mode = Standard, minimum length = 5 CV, width tolerance = 1 ms, centroiding mode = max, pad transitions = 5 CV). Independently, the same CSV files were processed with an in-house python script to calculate the $CIU50_{IWAT}$ as follows. Firstly, the intensity-weighted average arrival time (IWAT) was determined for each collision energy:

$$IWAT = \frac{\sum_{i=1}^{n} I_i t_i}{\sum_{i=1}^{n} I_i}$$

Equation S1

where *I* is the intensity and *t* is the arrival time of a given bin, *i,* in the mobility data. The intensity-weighted average arrival time or centroid time, was then fitted to a four-parameter sigmoid curve (using `scipy.optimise.curve_fit`) as in Equation S2, where *c* is the folded centroid time, (*a+c*) is the unfolded centroid time, *k* is the steepness of the transition and x_{50} is the $CIU50_{IWAT}$.

$$f(x) = \frac{a}{1+e^{-k(x-x_{50})}} + c$$

Equation S2

Both $CIU50_{FD}$ and $CIU50_{IWAT}$ represent the 50% or midpoint unfolding of the protein ions and depend on multiple individual measurements. Scripts for calculating $CIU50_{IWAT}$ are available at gist.github.com/jbellamycarter. Multiple replicate measurements of the wild-type ACP4 were used to determine a mean and standard deviation (σ) used for gauging significance of any (de)stabilisation effected by mutation to alanine. These were subtracted from the calculated midpoint values for each mutant to give $\Delta CIU50_{FD}$ and $\Delta CIU50_{IWAT}$ values, where the wild-type has a mean ΔCIU50 of zero and negative values indicate destabilisation of the protein in the gas-phase. The resulting values are shown in Table S5.

Calibration of TWIMS to CCS

The IMS–MS data from the travelling-wave IMS (TWIMS) of the Synapt was calibrated under the conditions shown in Table S2 with the Trap held at 5 V. Calibrant proteins used were: denatured

myoglobin (in 50% acetonitrile, 0.1% formic acid), native cytochrome c (25 mM AmAc) and native ubiquitin (25 mM AmAc), all proteins were 5 µM. ATDs were extracted and the peak maxima taken and compared to literature collisional cross section (CCS) values from the Clemmer[4] (for myoglobin) and Bush[5] (for cytochrome c[6] and ubiquitin[7]) CCS databases following the method outlined by Ruotolo *et al.*[8] The values used for this calibration are found in Table S3. The EDC coefficient was 1.41, this was used to correct drift times for *m/z* dependent delay. Calibration factors calculated were *X* = 0.328, *m* = 608.52 and *c* = 51.384, which were then used to correct PksACP4 drift times, see Figure S2.

Table S3 | Calibration of TWIMS drift times to CCS_{He}. See Figure S2 for fitting plots.

Protein	z	m/z	Drift Time (ms)	CCS_{He} (Å²)	$^{TW}CCS_{N2 \rightarrow He}$ (Å²)	Deviation
Myoglobin[4]	17	998.1	5.76	3384	3414	-0.9%
	16	1060.4	6.30	3313	3220	-0.2%
	15	1131.1	6.93	3230	3222	0.2%
	10	1696.1	4.59	1897	1870	1.4%
	9	1884.4	4.95	1758	1734	1.4%
	8	2119.9	5.40	1673	1595	4.7%
Cytochrome c[5,6]	7	1766.6	4.50	1280	1315	-2.7%
	6	2060.8	5.94	1240	1251	-0.9%
Ubiquitin[5,7]	6	1428.5	3.51	1000	1038	-3.8%
	5	1714.0	5.13	983	999	-1.7%
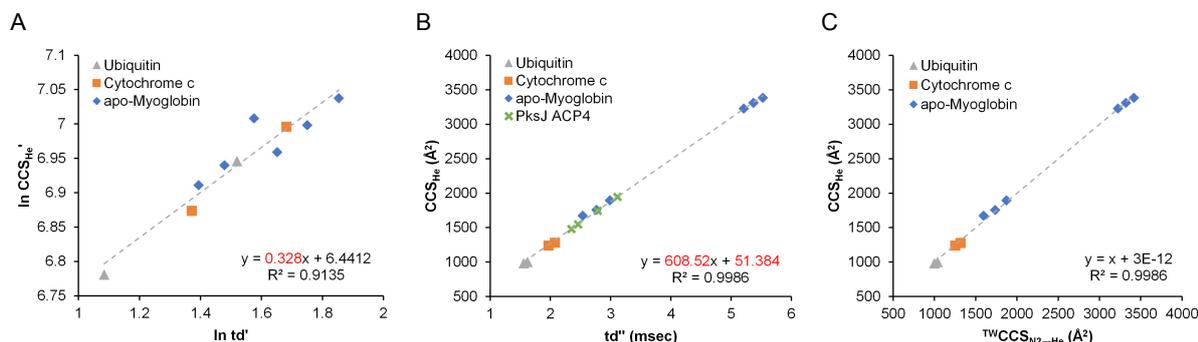
Figure S2 | Plots for calibration of TWIMS drift times to CCS_{He} with ubiquitin, cytochrome c and apo-myoglobin. A) Delay corrected drift time (ln t_d') against corrected literature CCS (ln CCS_{He}') giving exponential factor *X*=0.328. B) Exponential, charge and reduced mass corrected drift time (t_d'') against literature CCS_{He} giving coefficient *m*=608.52 and *c*=51.384. C) TWIMS calibrated CCS ($^{TW}CCS_{N2 \rightarrow He}$) against literature CCS_{He}.

Molecular Dynamics

All molecular dynamics experiments were performed on a computer running Ubuntu 16.04 and 18.04 LTS releases with GROMACS 5.1.2[9] installed from the apt package manager. Supplementary scripts were written in bash, batch, expect/tcl and python and are provided at gist.github.com/jbellamycarter and github.com/jbellamycarter/chargePlacer.

Charge Placement Algorithm

In order to make charge assignments tractable and reproducible, a python tool (ChargePlacer) was developed to determine energy minimised proton sequences for a given input structure. Proton

sequences were used as an abstraction from the charge sequence as this simplified the rearrangement to a binary form where the number of protons was fixed, greatly increasing the speed of calculation.
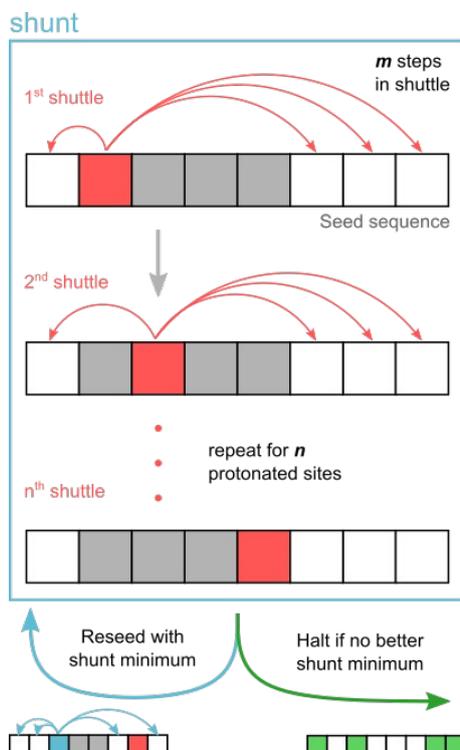
Figure S3 | Schematic representation of the proton distribution algorithm.
The algorithm is broken into nested stages. 1) Shuttle, *m* single-proton steps for *m* unoccupied sites and 2) Shunt, *n* shuttles for *n* unoccupied sites.

A visual depiction of the algorithm is given in Figure S3. Briefly, a randomised proton sequence is taken as the seed; for a given set of chargeable sites (*s*) and target charge, there are *n* protons and *m* unoccupied sites (such that the sum of the deprotonated charges and *n* is the target charge; and that *m* + *n* = *s*). A single proton is then moved to each of the unoccupied sites and the energy of that sequence calculated (see Energy Calculation), this is performed *m* times and is called a *Shuttle*. A shuttle is performed for each proton (*n* times), giving a total of *m×n* sequences, this is called a *Shunt*. The sequence with the lowest energy in the shunt is carried forward (shunt minimum). If shunt minimum < current minimum energy, then this sequence is used to reseed the algorithm; if the shunt minimum ≥ current minimum, the algorithm exits and returns the sequence which gave the current minimum.

This algorithm may sample the same sequence multiple times but generally converges well before sampling the whole combination space. For ACP4^{8+} there are 40 chargeable sites requiring 29 protons, ChargePlacer converges within 9 shunts (2871 steps), see Table S6 for details. Whereas the total number of combinations *(n + m)! / (m! × n!)*, being 2.31×10^9, would require 20 hours of computation time on the same computer.

To test the robustness of the algorithm, ChargePlacer was run 100 times with randomised seed sequences on ACP4 (8+, see Figure S49) and Lysozyme (8+, 1AKI, see Figure S50), and 10 times on the multi-chain protein transthyretin, TTR (14+, 3GRG, see Figure S51) structures. These showed that for

all proteins tested, the algorithm converged to consistent proton patterns. These also showed that, generally, the Coulomb-only mode generated charge patterns with greater zwitterionic character.

An additional feature of ChargePlacer is the *in silico* alanine scanning function, an example of this applied to ACP4[8+] can be seen in Figure S52 and Table S7. This iterates through each mutable chargeable site in the structure, calculating the minimised proton pattern for the structure in the absence of that site. With the speed of the energy minimisation algorithm, this can be performed rapidly on small proteins. For larger proteins with multiple chains, chains can be 'protected' from the scan to improve analysis time.

Energy Calculation

ChargePlacer uses a point-charge approximation for calculating the energies to minimise. The point charges are assigned per chargeable site: Asp, Glu, Lys, Arg, His, N-terminus and C-terminus; see Table S4 for details.

Table S4 | Point-charge details for ChargePlacer energy calculation.

Residue	Asp (D)	Glu (E)	Lys (K)	Arg (R)	His (H)	N-ter (NT)	C-ter (CT)
Point-charge atom (OPLS-AA)[10]	OD2	OE2	NZ	NH2	CB	N	C
Deprotonated charge	-1	-1	0	0	0	0	-1
Proton affinity (kJ/mol)[11]	1453.5	1448.5	918	1002	958	886.6	1430

A single atom per site is used (in OPLS-AA/PyMOL notation), being the most likely site for protonation. For a given proton sequence, the charge sequence is trivially determined by adding the proton sequence to a fixed deprotonated charge sequence. When the algorithm is initialised for a given structure this deprotonated charge sequence is set along with a distance matrix (for each point–point interaction (r_{ij}), these are calculated only once and reused for all subsequent calculations. The Coulomb energy (E_{Coul}) is calculated with Equation S3, where e is the elementary charge, ε_0 is the vacuum permittivity, ε_r is the relative permittivity of the medium, i and j are the two point-charge atoms, z is the charge of the atom and r_{ij} is the distance between the atoms.

$$E_{Coul} = \frac{e^2}{4\pi\varepsilon_0\varepsilon_r}\sum_{i=1}^{n}\sum_{j=i+1}^{n}\frac{z_i z_j}{r_{ij}}$$

Equation S3

The energy calculation can be set to either account for E_{Coul} only, or to account for the energy released upon binding of a proton (see Table S4 for proton affinities used). If proton affinities are accounted for (the default setting in ChargePlacer), the total energy (E_{tot}) is calculated with Equation S4, where the proton affinities (*PA*) for all protonated residues are summed.

$$E_{tot} = E_{Coul} - \sum_{k=1}^{n} PA_k$$

Equation S4

We have generally observed that using the Coulomb-only setting produces charge sequences with more zwitterionic character, and that this results in slightly more stable trajectories over thermal gradient MD simulations for PksJ ACP4 (see Figure S53).

Gas-Phase Simulations of ACP4

The placement of protons to reach the 8+ charge state for ACP4 was performed using the iTASSER structure with ChargePlacer. The resulting proton patterns (in the `proton_sites.txt` file) were used to generate the appropriately charged input gro file using an `expect/tcl` script (`autocharger.exp`) that runs a `gmx-pdb2gmx` child process (shell script `pdb2gmx_automate.sh`) with the OPLS-AA forcefield. To allow the use of the efficient Verlet cut-off scheme, the protein was then centred in a very large cubic boundary box (900×900×900 nm) and energy minimised by steepest descent for 10,000 steps (`min.mdp`). For all gas-phase simulations, long-range interactions (Coulomb and Van der Waals) were set to be extremely large (300 nm) in a periodic boundary box. These are required to simulate gas-phase interactions accurately with the Verlet scheme, as discussed by Konermann.[12]

The energy-minimised protein structures were then equilibrated at 298 K for 50 ps with a velocity rescaling thermocouple, a pseudorandom number generator was used to seed the velocities. H-bonds were constrained using the LINCS algorithm (iter=1, order=4). For the production run, this trajectory was continued for 1 ns. For thermal gradient runs, annealing was performed from 298 K to 798 K linearly over the 1 ns. Further 4 ns continuations were performed using thermal gradient end-points, these were maintained at 798 K to simulate the fully unfolded protein ions. A step-size of 2 fs was used for all simulations.

For the production run simulations, built-in GROMACS[9] analysis functions were used to calculate the solvent-accessible surface area (`gmx-sasa`), radius of gyration (`gmx-gyrate`) and root-mean-squared deviation (`gmx-rms`) of trajectories. To compare with CIU data, some simulations were converted to PDB files (`gmx-trjconv`) and then modelled with the projection approximation (PA) model using CCSCalc software (Waters) with a gas radius of 1.4 and a modified atom types file (`types_oplsaa.txt`)—modified to include OPLS-AA atom names. The resulting CCS_{PA} values were multiplied by the empirically determined 1.14 to give CCS_{calc}, as the PA model is known to underestimate CCS values.[13] These CCS_{calc} values were binned at bin width 10 Å² and compared to $^{TW}CCS_{N2 \rightarrow He}$ corrected PksACP4 IMS data (see Calibration of TWIMS to CCS).

Solution-Phase Simulations of ACP4

The initial structure file was generated using the iTASSER model from above with `gmx-pdb2gmx` using the OPLS-AA/L forcefield[10] and TIP3P water model. A triclinic box (1 nm) was used, filled with water (`gmx-solvate ↠ -cs spc216.gro`) and sodium counter-ions (`gmx-genion`) introduced to neutralise the box. The structure was energy minimised by steepest descent for 50,000 steps with PME. An NVT run was performed for 100 ps at 298 K (velocity rescaling thermocouple) with positional restraints. A subsequent NPT run was performed for 100 ps at 298 K (Parrinello-Rahman pressure coupling). The production run was performed for 1 ns at 298 K. As above, H-bonds were constrained using the LINCS algorithm (iter=1, order=4) and velocities generated with a pseudorandom seed number.

Bioinformatic Analysis of *trans*-AT PKS ACPs

Some 315 ACPs from 21 *trans*-AT PKS and 1 *cis*-AT PKS were collated manually from MiBIG[14] and NCBI databases. These were grouped by the clade of their up/down-stream KS before aligning sequences with Clustal Omega[15,16] and generating hidden Markov Model (HMM)[17,18] profiles for each group. Any KS without predefined clades were classified using BLAST according to Nguyen et al.

[19] Sequence logos for these ACPs grouped by their downstream KS-clade are shown in Figure S57. These HMM profiles were then compared with the HMM profile for the whole collection, in the same manner as Haines et al.,[20] allowing the validity of grouping the ACPs in such a way. Figures S58–S60 show that grouping by downstream KS-clade is highly discriminative, suggesting that sequences are conserved for common ACP–KS pairs. Furthermore, grouping by the nearest enzymatic domain upstream of the ACP is also a good discriminator.

The PKS pathways used were: bacillaene (Bae, BGC0001089, n=19), batumin (Bat, BGC0001099, n=16), bongkrekic acid (Bon, BGC0000173, n=13), bryostatin (Bry, BGC0000174, n=18), chivosazol (Chi, BGC0001069, n=24), diaphorin (Dip, BGC0001092, n=13), difficidin (Dif, BGC0000176, n=18), disorazol (Dsz, BGC0001093, n=11), enacyloxin (Ena, BGC0001094, n=12), kirromycin (Kir, BGC0001070, n=17), lankacidin (Lkc, BGC0001100, n=5), leinamycin (Lmn, BGC0001101, n=10), macrolactin (Mln, BGC0001383, n=15), mupirocin (Mmp, BGC0000182, n=11), myxovirescin (Ta, BGC0001025, n=17), nosperin (Nsp, BGC0001071, n=14), onnamide (Onn, BGC0001105, n=11), pederin (Ped, BGC0001108, n=17), psymberin (Psy, BGC0001110, n=13), rhizoxin (Rhi, BGC0001112, n=20), thailandamide (Tai, BGC0000186, n=20) and erythromycin (Ery, BGC0000055, n=1 (ACP4)).

Supplementary Data
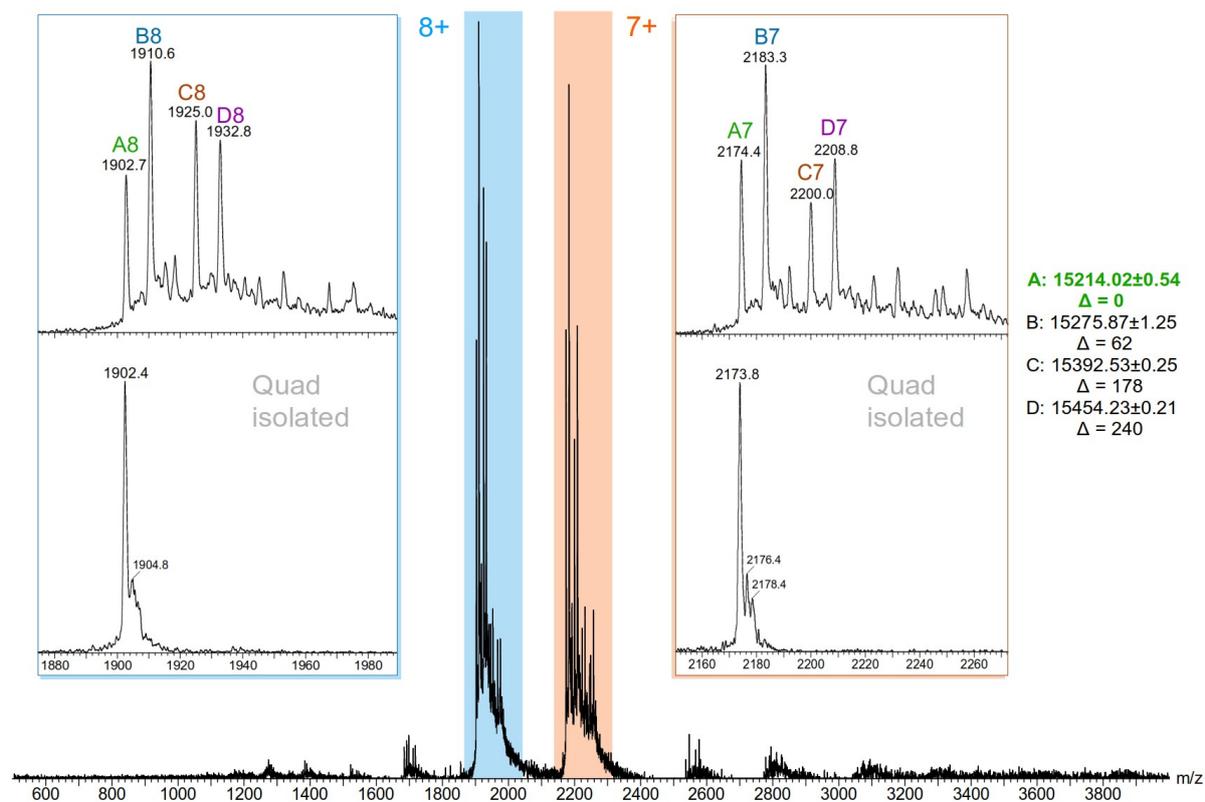
Figure S4 | Native mass spectra of ACP4 showing the heterogeneous species observed.
A) expected unmodified ACP4 without N-terminus, B) acetate adduct of A, C) N-gluconylation of A, and D) acetate adduct of C. Insets show the zoomed in regions for the 8+ (*left*) and 7+ (*right*) charge states and results of quad isolation of the unmodified protein species (A). Quad isolation sufficiently removes modified species and adducts. Additional buffer exchange did not affect the composition of species.
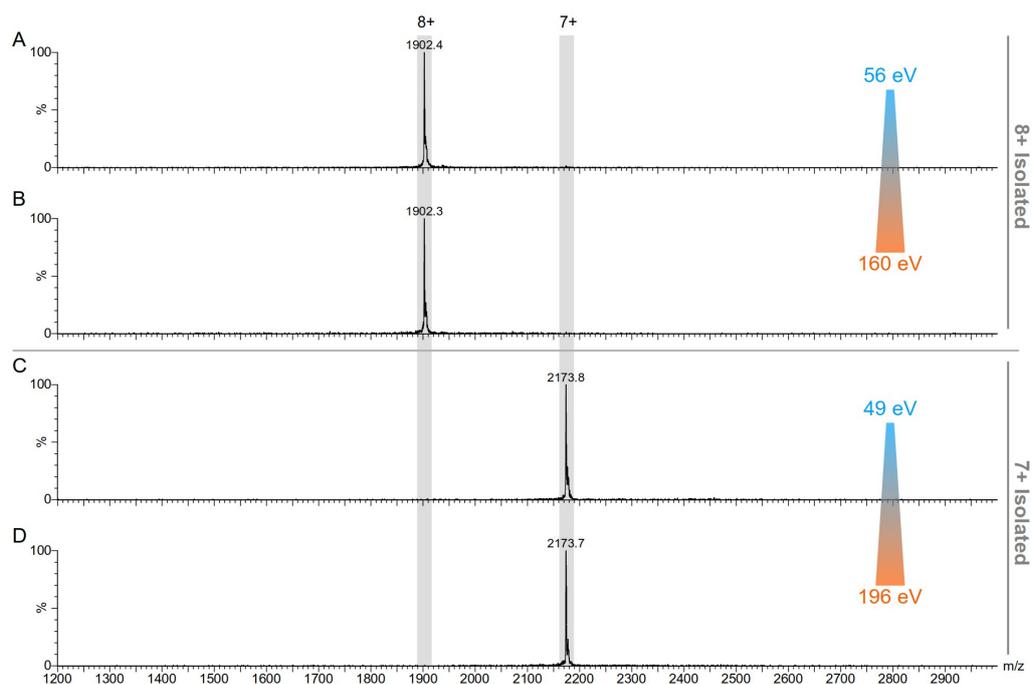
Figure S5 | Quadrupole isolated mass spectra of the wild-type ACP4.
A and B) 8+ charge state. C and D) 7+ charge state. At low (A and C) and high (B and D) energies showing no charge stripping or fragmentation over the full range of activation energies used.

CIU Data

Plots of CIU data collected for all ACP4 variants for both 8+ and 7+ charge states. For each variant (or replicate) a six-panel figure is given containing the CIU heat-maps (A and D), CIU50$_{FD}$ fitting from CIUSuite2 (B and E) and CIU50$_{IWAT}$ fitting from in-house scripts (C and F) for the 8+ (A, B and C) and 7+ (D, E and F) charge states.

Wild-type PksJ ACP4

Several replicates of the wild-type protein were acquired over several months, to act as a baseline in the event of instrumental drift. Replicate 1 is the base used for data collected between October and December 2019, while replicates 2–6 were used as the base for data collected between January and March 2020. Some significant instrumental drift was observed between these two time periods, which we principally attribute to instrument maintenance performed in December. Use of the multiple wild-type standard measurements allowed assessment and correction of these effects. The error estimation for CIU50 was determined as the standard deviation of replicates 2–6, which were acquired independently over 5 weeks (10th Feb – 17th Mar).
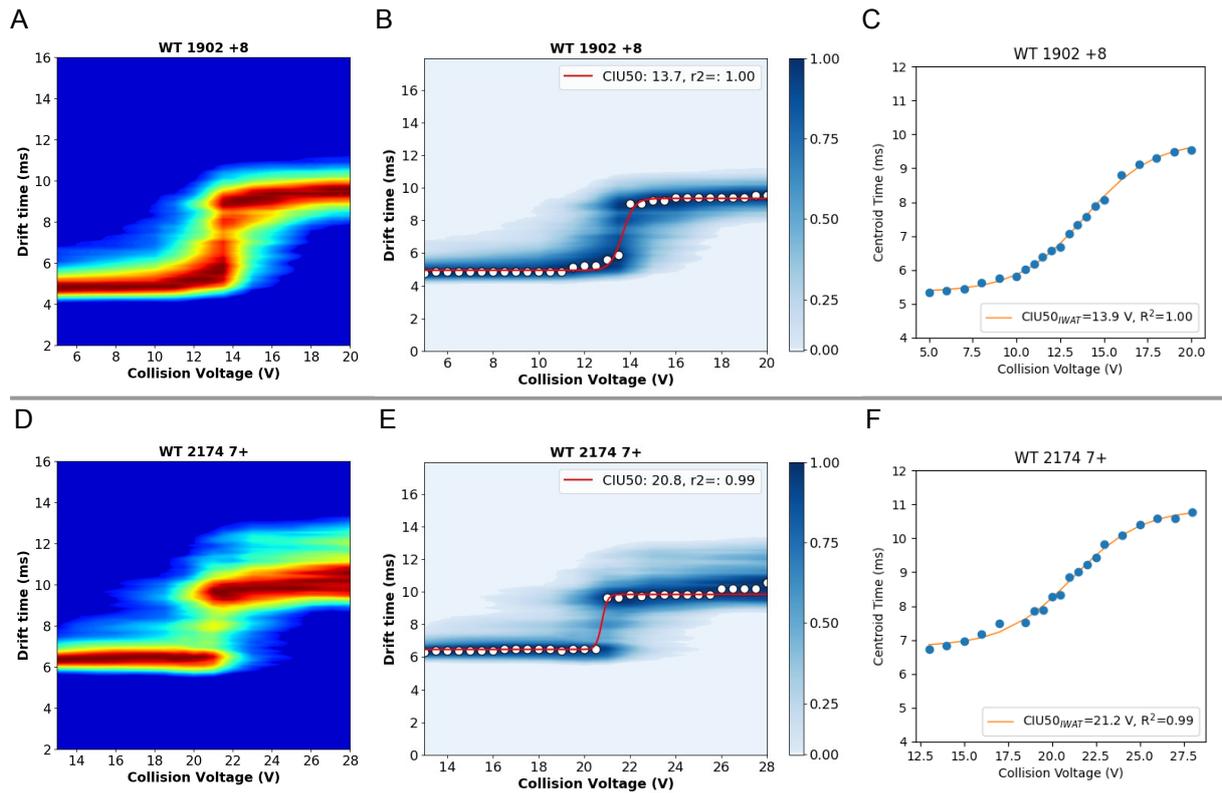
Figure S6 | CIU data for **WT** replicate **1** of PksACP4 for the 8+ (A, B and C) and 7+ (D, E and F) charge states. A and D) CIU plot for the 8+ charge state. B and E) CIU50$_{FD}$ fitting from CIUSuite2. C and F) CIU50$_{IWAT}$ fitting from in-house scripts.
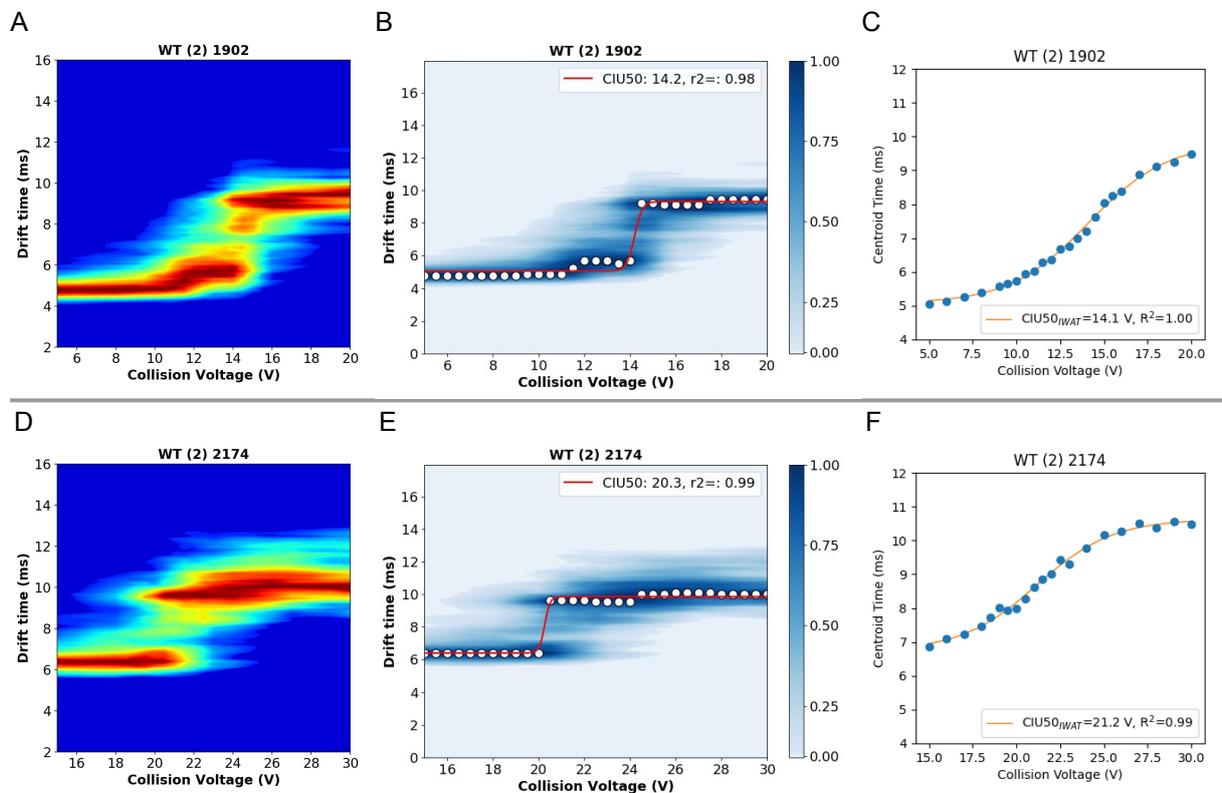
Figure S7 | CIU data for **WT** replicate **2** of PksACP4 for the 8+ (A, B and C) and 7+ (D, E and F) charge states. A and D) CIU plot for the 8+ charge state. B and E) CIU50$_{FD}$ fitting from CIUSuite2. C and F) CIU50$_{IWAT}$ fitting from in-house scripts.
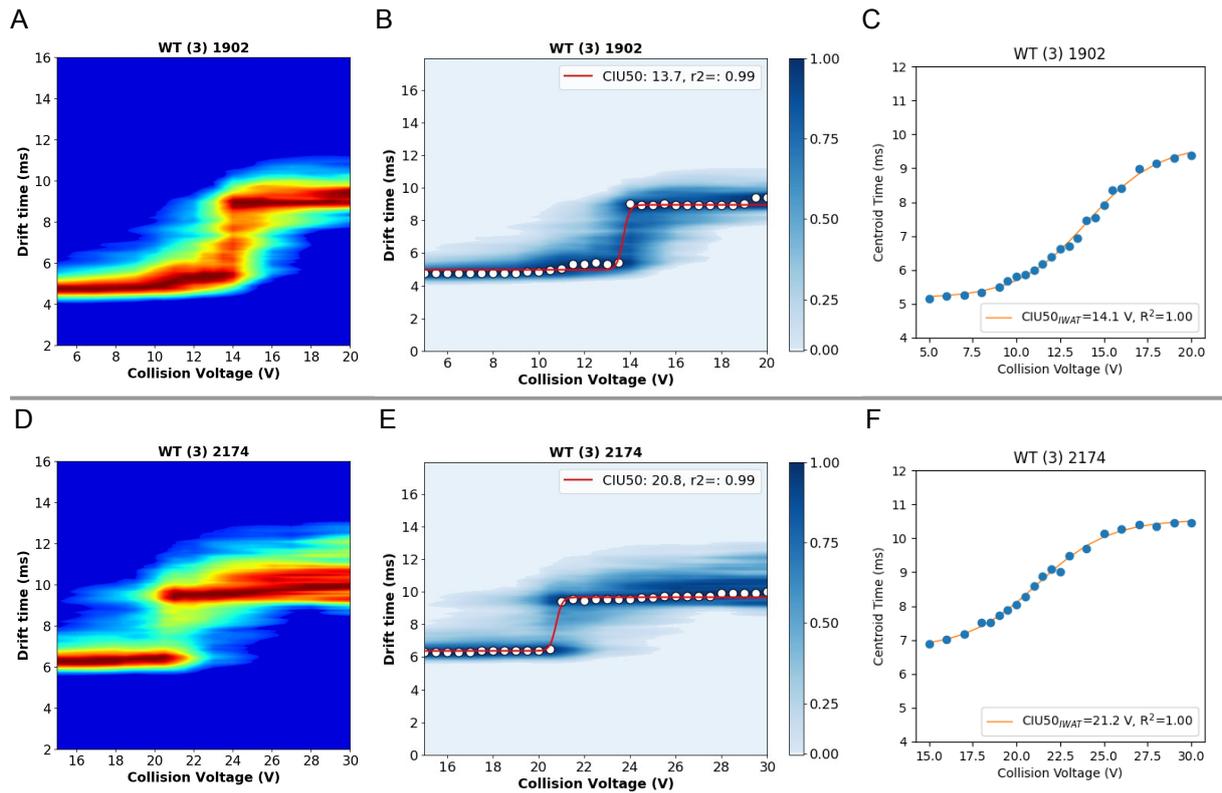
Figure S8 | CIU data for **WT** replicate **3** of PksACP4 for the 8+ (A, B and C) and 7+ (D, E and F) charge states. A and D) CIU plot for the 8+ charge state. B and E) CIU50$_{FD}$ fitting from CIUSuite2. C and F) CIU50$_{IWAT}$ fitting from in-house scripts.
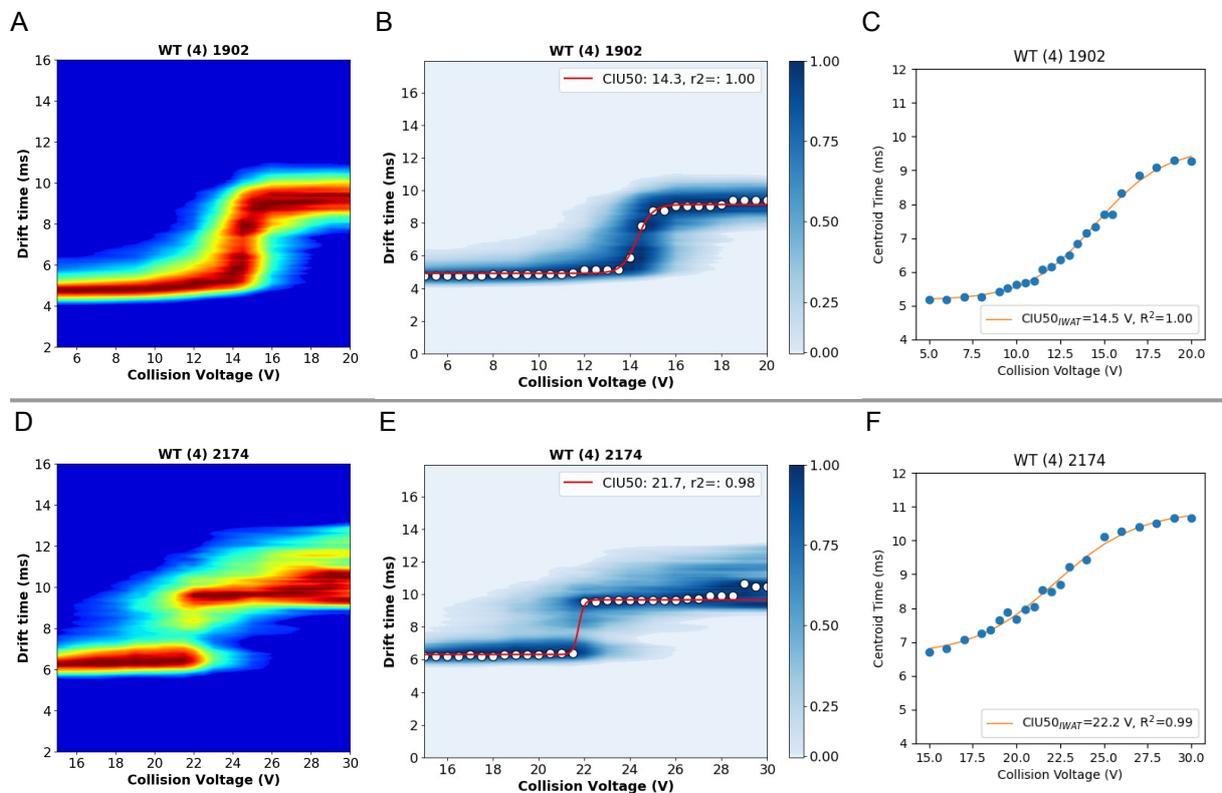
Figure S9 | CIU data for **WT** replicate **4** of PksACP4 for the 8+ (A, B and C) and 7+ (D, E and F) charge states. A and D) CIU plot for the 8+ charge state. B and E) CIU50$_{FD}$ fitting from CIUSuite2. C and F) CIU50$_{IWAT}$ fitting from in-house scripts.
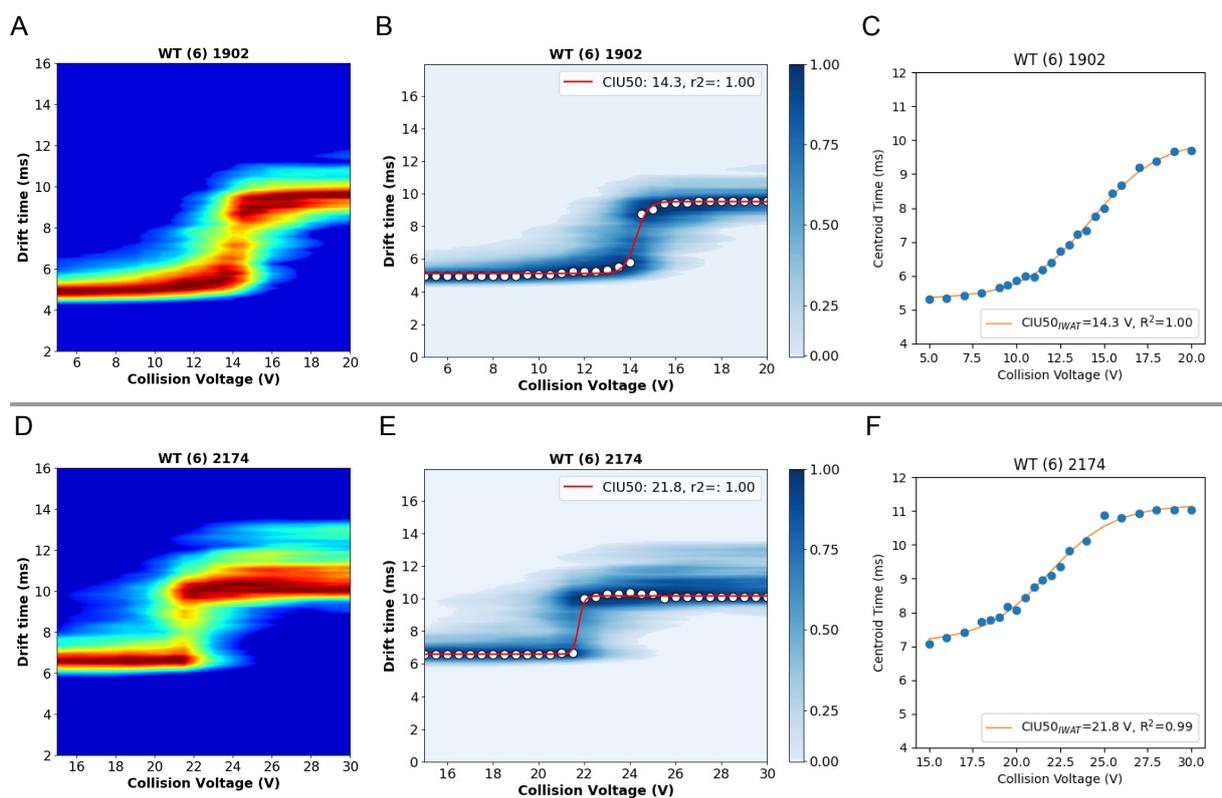
Figure S10 | CIU data for **WT** replicate **6** of PksACP4 for the 8+ (A, B and C) and 7+ (D, E and F) charge states. A and D) CIU plot for the 8+ charge state. B and E) CIU50$_{FD}$ fitting from CIUSuite2. C and F) CIU50$_{IWAT}$ fitting from in-house scripts.

Mutants

The mutant data have been ordered according to residue type. A summary of CIU50 values recorded for each mutant is given in Table S5. For mutants D41A and K59A, the three replicates used for t-test and the mean CIU50 are shown in Figures 15, 16, 17 and 29, 30, 31, respectively.

Table S5 | Determined ΔCIU50 values for PksJ ACP4 mutants compared to wild-type. Theoretical molecular masses were determined using ExPASy ProtParam. [a]Mutants recorded in triplicate.

Type	Residue	Theoretical Values			$\Delta CIU50_{IWAT}$ (eV)		$\Delta CIU50_{FD}$ (eV)	
		Mass (Da)	m/z (8+)	m/z (7+)	8+	7+	8+	7+
Ser (S)	67	15200.92	1901.12	2172.56	-0.5	1.2	0.6	4.7
Pro (P)	66	15190.88	1899.86	2171.13	1	1.8	-0.1	7.9
Leu (L)	26	15174.84	1897.86	2168.83	-0.7	-2.1	0.8	1.4
Ile (I)	28	15174.84	1897.86	2168.83	-2.3	3.2	-1.2	4.2
Asp (D)	20	15172.91	1897.61	2168.56	-5.6	-4.7	-2.1	-6
	29				-5	-3.6	-2.8	-5.9
	32				-5.1	-11.6	-3.4	-5.5
	36				-1	-2.5	-3.8	-2.4
	41[a]				-6 ± 0.23	-7.5 ± 0.99	-7.6 ± 0.73	-7.3 ± 2.02
	65				-9.4	-9	-7.4	-6.2
	80				0.3	-1.1	0.4	1.1
Gln (Q)	15	15159.87	1895.98	2166.7	1.4	0.5	1	-2.7
	40				-1.4	-0.5	0.5	0.8
	51				0.6	-1.5	0.2	0.9
	54				-1.3	-1.9	-0.5	0.9
	76				-0.3	-0.7	0.1	4.3
Lys (K)	59[a]	15159.82	1895.98	2166.69	-3.9 ± 0.55	-3.8 ± 0.40	-2.6 ± 0.08	-3.7 ± 1.96
Glu (E)	13	15158.88	1895.86	2166.55	-3.2	-1.3	-3.7	-0.1
	24				-1.8	-3.2	1.2	0.5
	25				-4.5	-6.4	-4.4	-3.6
	31				-5.3	-8	-6.9	-3.4
	34				-6.4	-2.7	-4.7	-2.7
	61				-1	-3.4	-3	-7.1
	71				-3	-2.9	-5.8	-6.9
Phe (F)	33	15140.82	1893.6	2163.97	1.8	-0.3	3.6	4.4
Arg (R)	27	15131.81	1892.48	2162.69	-3.6	-5.4	-3.3	-3.2
	30				-3.1	-6.4	-3.7	-2.7
	55				-3.3	-4.2	-3.4	1.2
	58				0.7	-1.3	0.5	1.2
	77				-4.2	-5.7	-4.6	-3.7
Tyr (Y)	42	15124.82	1891.6	2161.69	-0.6	-1.7	0	1
	70				-1.2	-1.3	1.1	1.1
	72				-0.4	-3	0.7	0.7
Trp (W)	81	15101.78	1888.72	2158.4	-3.4	-3.6	-3.3	-2.8
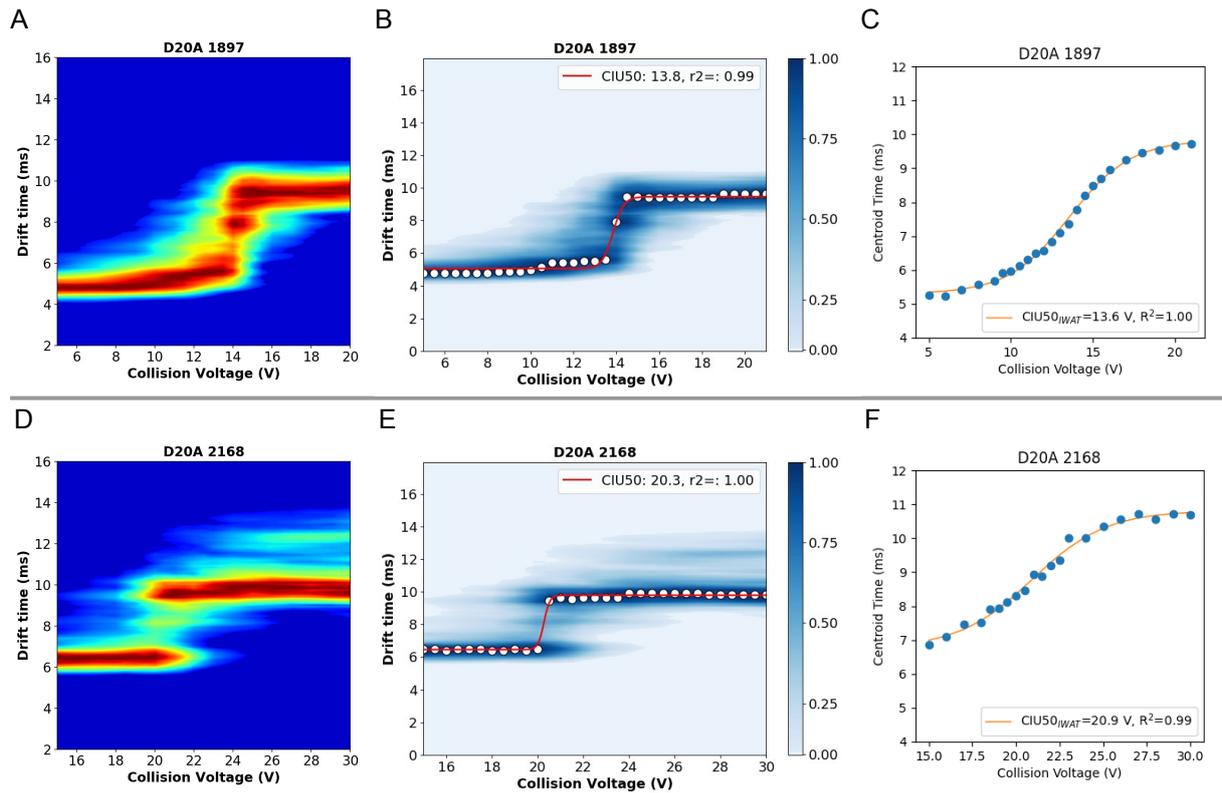
Figure S11 | CIU data for **D20A** mutant of PksACP4 for the 8+ (A, B and C) and 7+ (D, E and F) charge states. A and D) CIU plot for the 8+ charge state. B and E) CIU50$_{FD}$ fitting from CIUSuite2. C and F) CIU50$_{IWAT}$ fitting from in-house scripts.
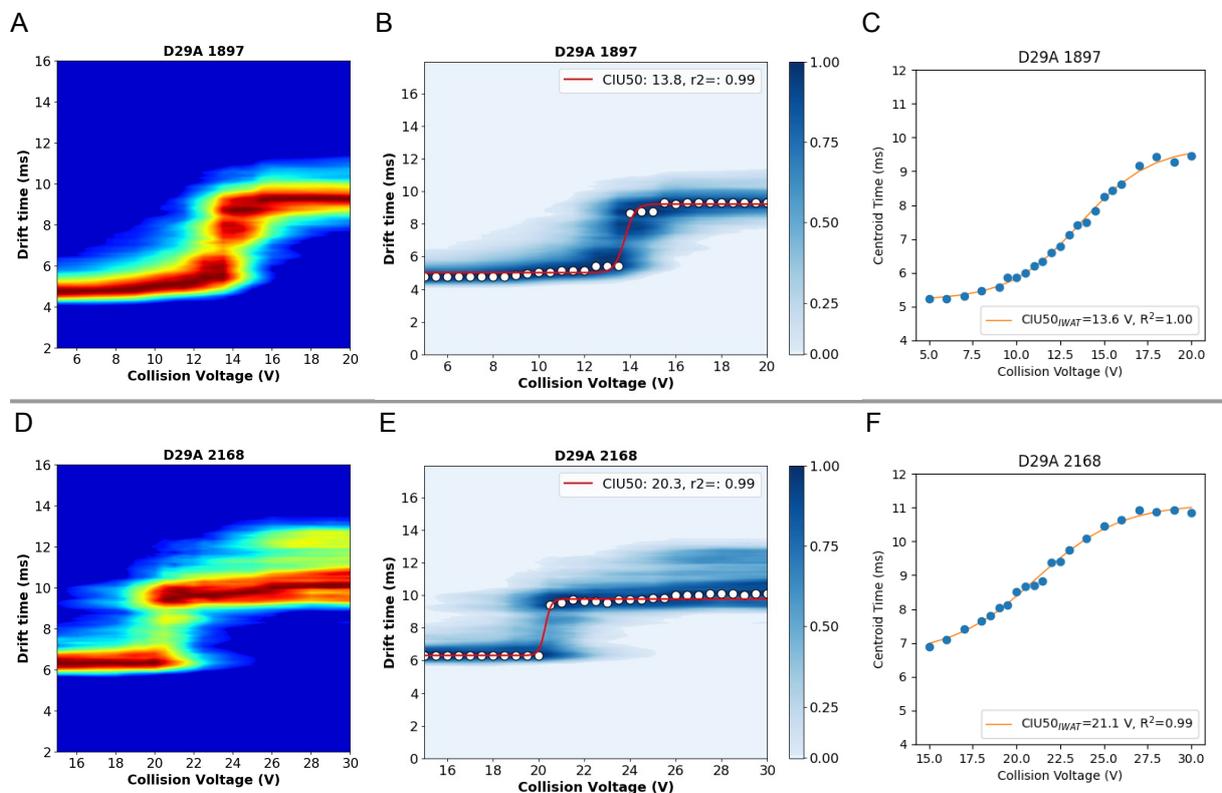
Figure S12 | CIU data for **D29A** mutant of PksACP4 for the 8+ (A, B and C) and 7+ (D, E and F) charge states. A and D) CIU plot for the 8+ charge state. B and E) CIU50$_{FD}$ fitting from CIUSuite2. C and F) CIU50$_{IWAT}$ fitting from in-house scripts.
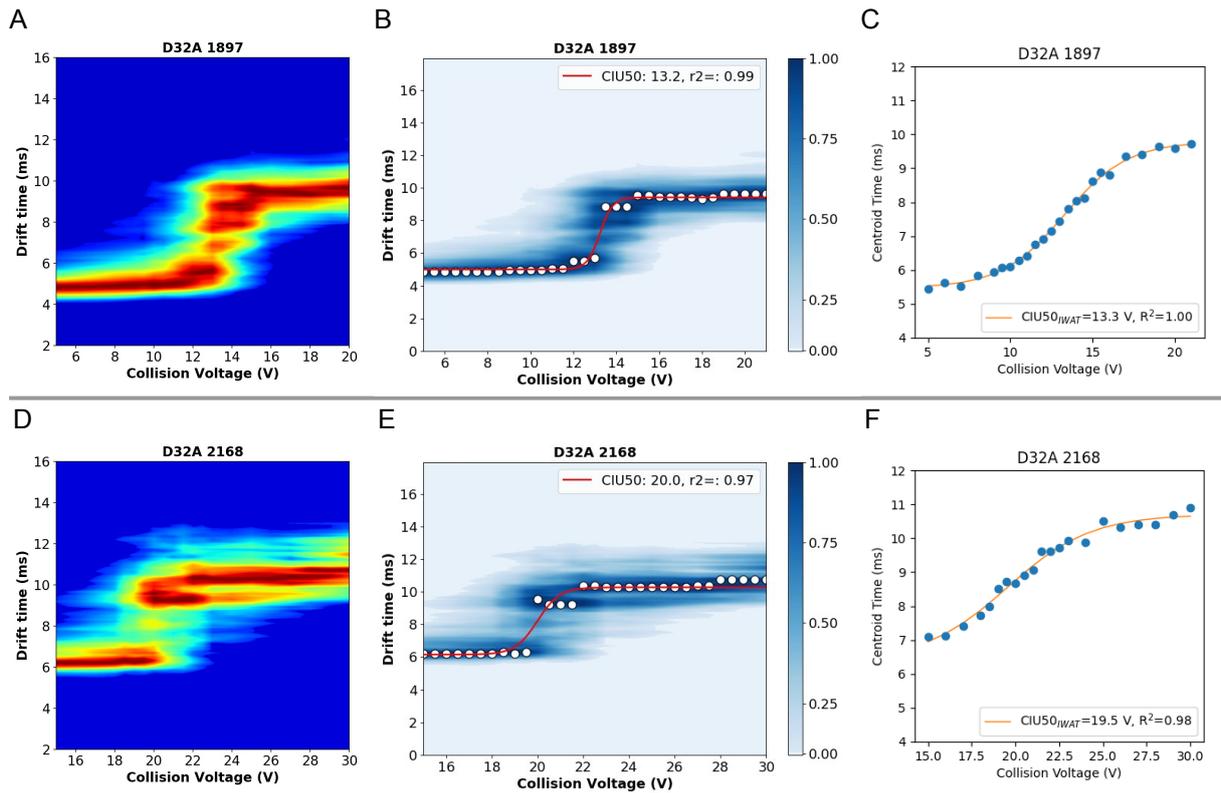
Figure S13 | CIU data for **D32A** mutant of PksACP4 for the 8+ (A, B and C) and 7+ (D, E and F) charge states. A and D) CIU plot for the 8+ charge state. B and E) CIU50$_{FD}$ fitting from CIUSuite2. C and F) CIU50$_{IWAT}$ fitting from in-house scripts.
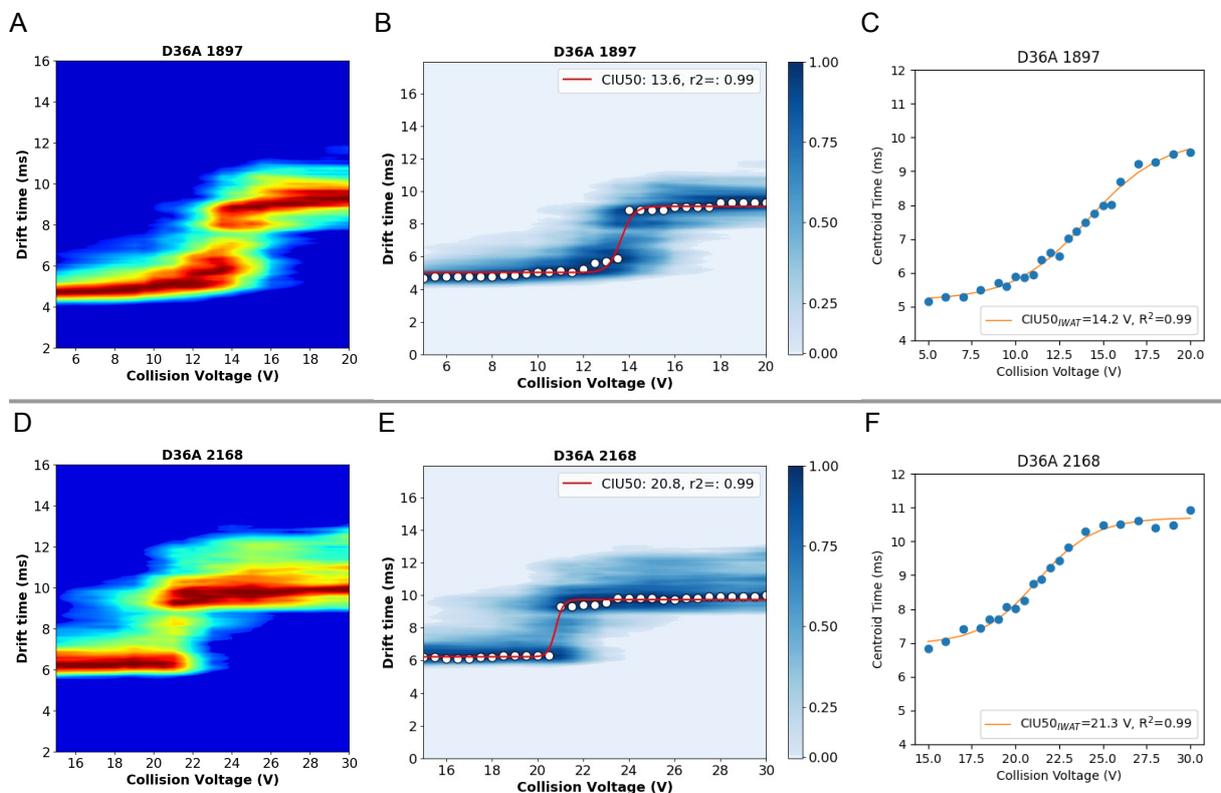
Figure S14 | CIU data for **D36A** mutant of PksACP4 for the 8+ (A, B and C) and 7+ (D, E and F) charge states. A and D) CIU plot for the 8+ charge state. B and E) CIU50$_{FD}$ fitting from CIUSuite2. C and F) CIU50$_{IWAT}$ fitting from in-house scripts.
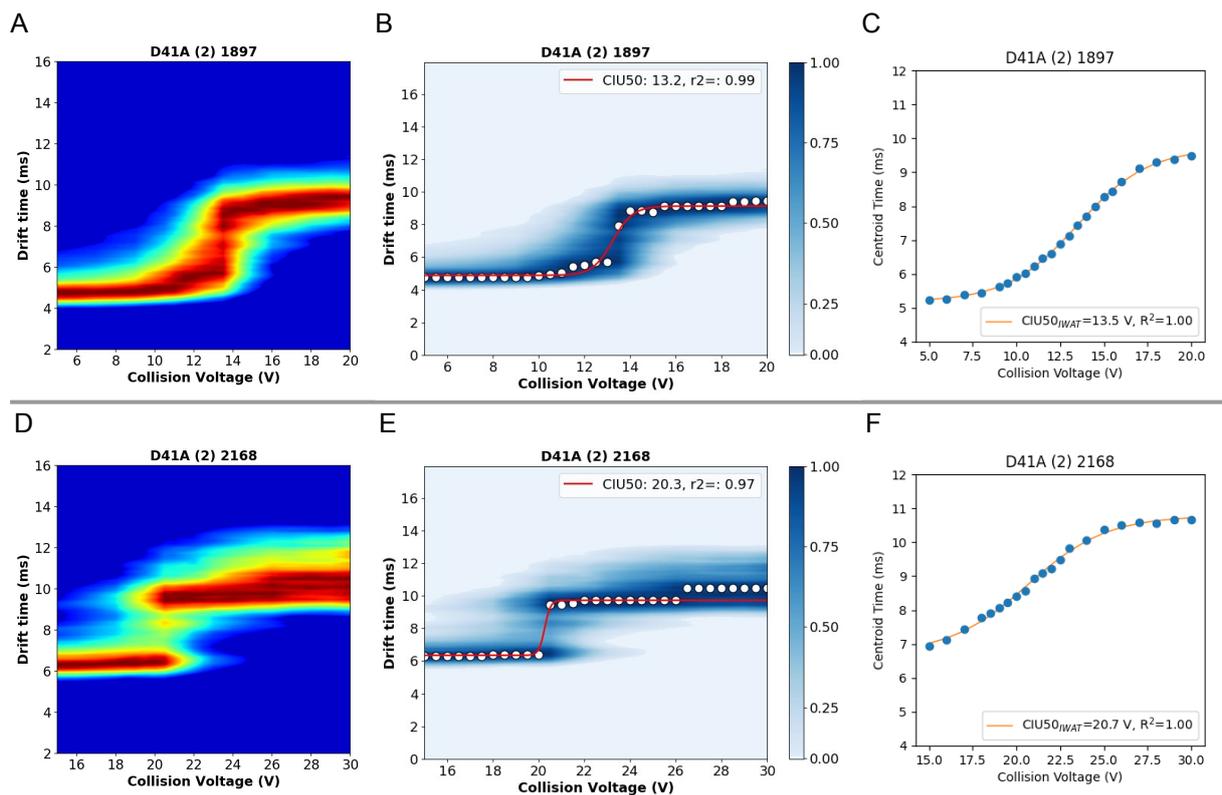
Figure S15 | CIU data for **D41A** mutant replicate **2** of PksACP4 for the 8+ (A, B and C) and 7+ (D, E and F) charge states. A and D) CIU plot for the 8+ charge state. B and E) CIU50$_{FD}$ fitting from CIUSuite2. C and F) CIU50$_{IWAT}$ fitting from in-house scripts.
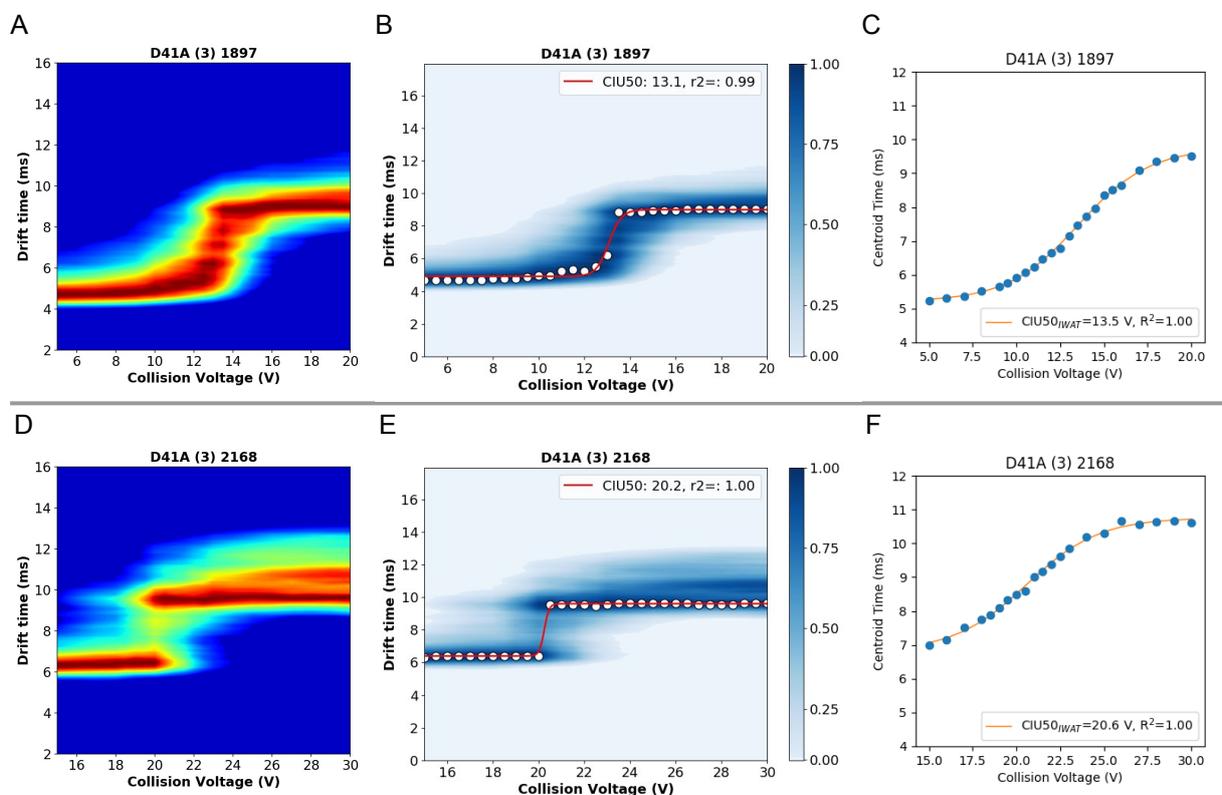
Figure S16 | CIU data for **D41A** mutant replicate **3** of PksACP4 for the 8+ (A, B and C) and 7+ (D, E and F) charge states. A and D) CIU plot for the 8+ charge state. B and E) CIU50$_{FD}$ fitting from CIUSuite2. C and F) CIU50$_{IWAT}$ fitting from in-house scripts.
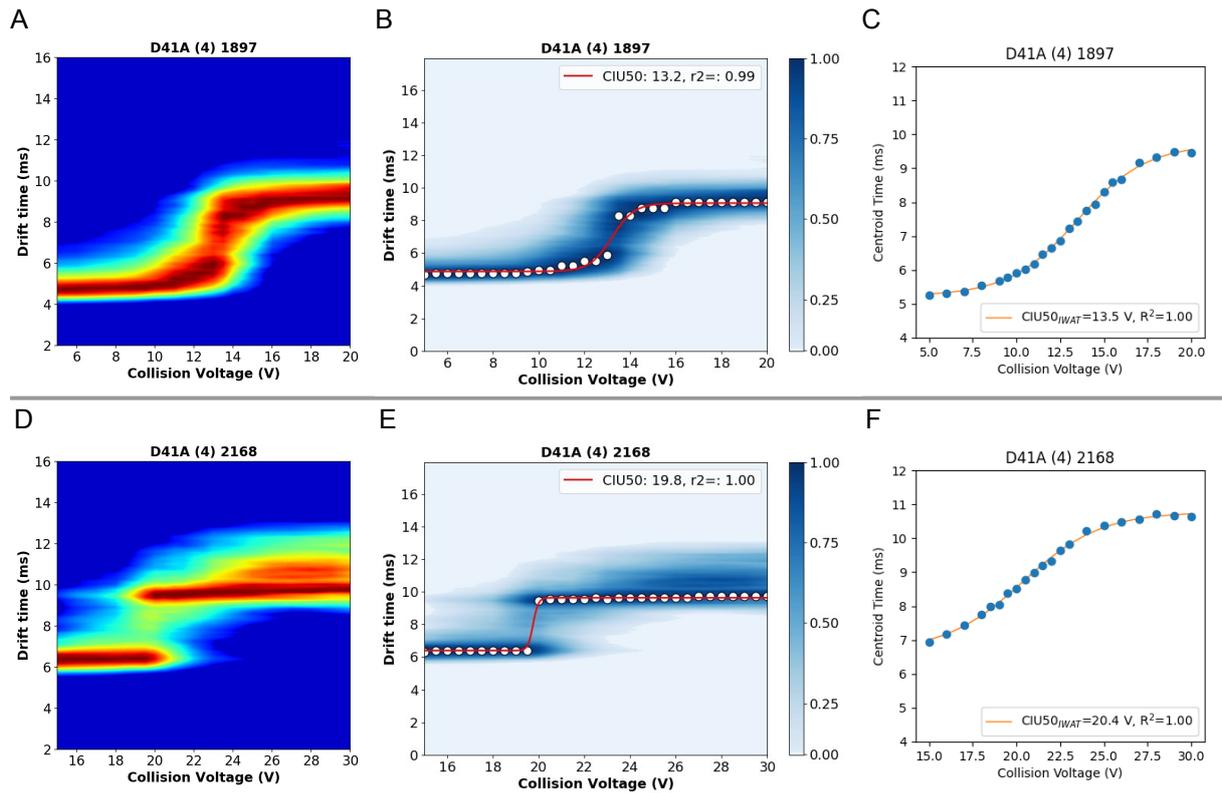
Figure S17 | CIU data for **D41A** mutant replicate **4** of PksACP4 for the 8+ (A, B and C) and 7+ (D, E and F) charge states. A and D) CIU plot for the 8+ charge state. B and E) CIU50$_{FD}$ fitting from CIUSuite2. C and F) CIU50$_{IWAT}$ fitting from in-house scripts.
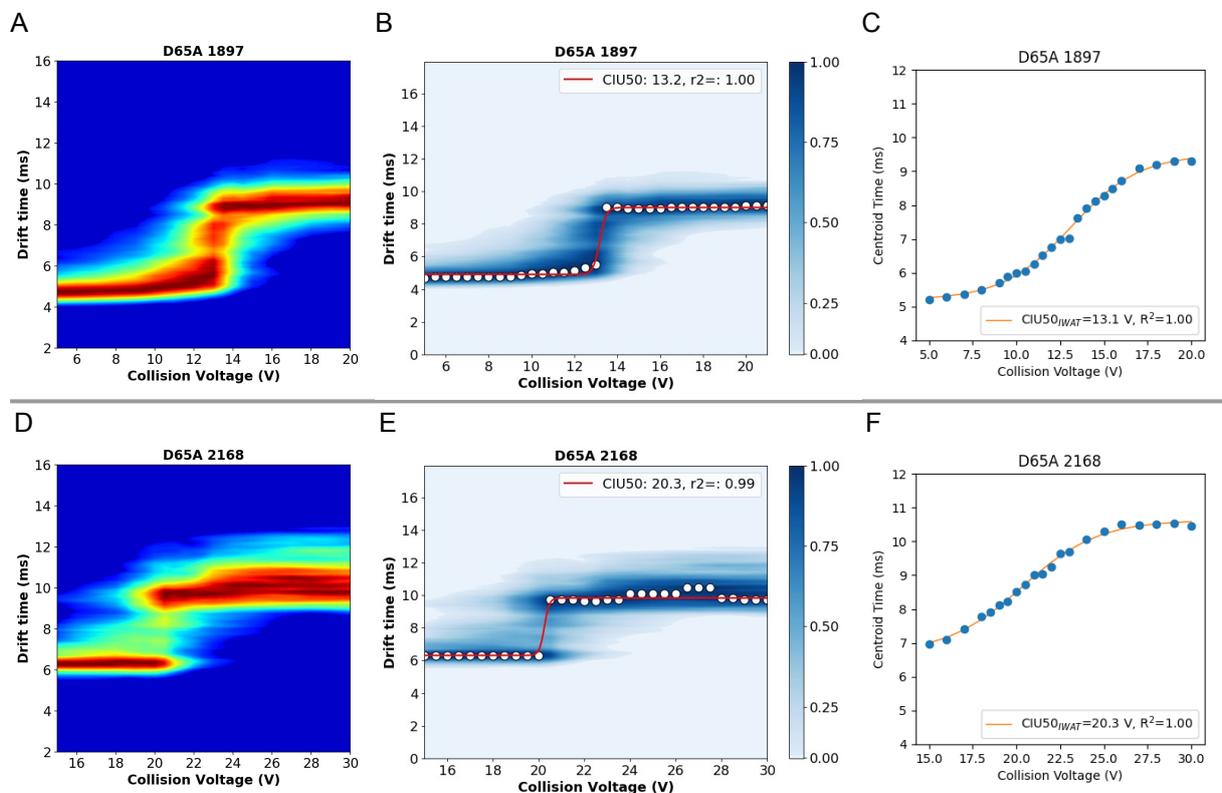
Figure S18 | CIU data for **D65A** mutant of PksACP4 for the 8+ (A, B and C) and 7+ (D, E and F) charge states. A and D) CIU plot for the 8+ charge state. B and E) CIU50$_{FD}$ fitting from CIUSuite2. C and F) CIU50$_{IWAT}$ fitting from in-house scripts.
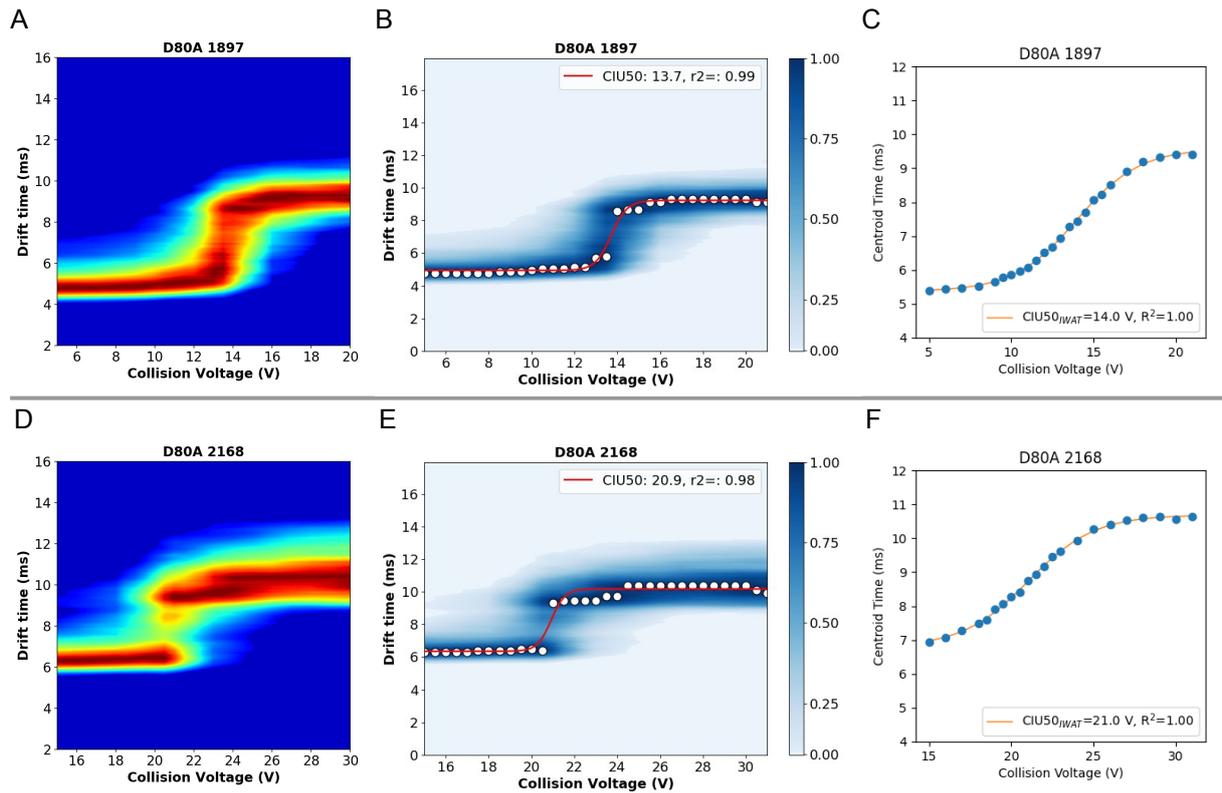
Figure S19 | CIU data for **D80A** mutant of PksACP4 for the 8+ (A, B and C) and 7+ (D, E and F) charge states. A and D) CIU plot for the 8+ charge state. B and E) CIU50$_{FD}$ fitting from CIUSuite2. C and F) CIU50$_{IWAT}$ fitting from in-house scripts.
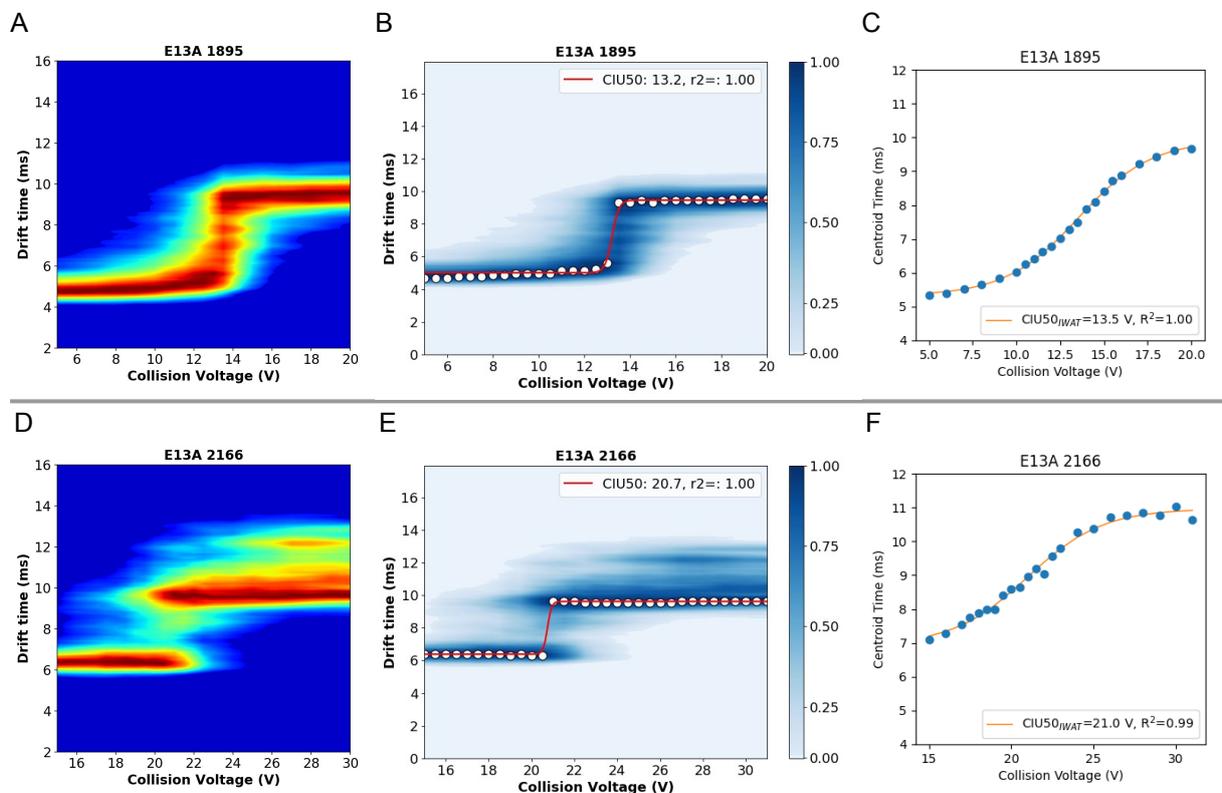
Figure S20 | CIU data for **E13A** mutant of PksACP4 for the 8+ (A, B and C) and 7+ (D, E and F) charge states. A and D) CIU plot for the 8+ charge state. B and E) CIU50$_{FD}$ fitting from CIUSuite2. C and F) CIU50$_{IWAT}$ fitting from in-house scripts.
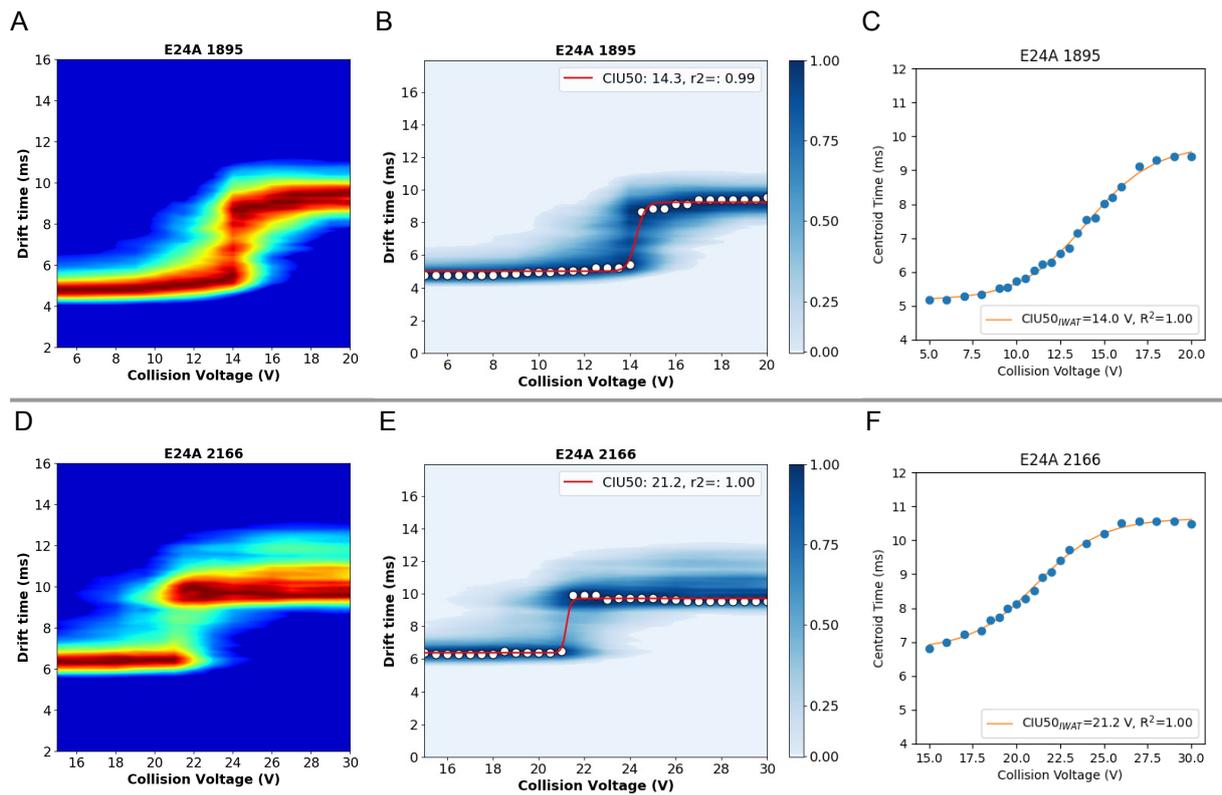
Figure S21 | CIU data for **E24A** mutant of PksACP4 for the 8+ (A, B and C) and 7+ (D, E and F) charge states. A and D) CIU plot for the 8+ charge state. B and E) CIU50$_{FD}$ fitting from CIUSuite2. C and F) CIU50$_{IWAT}$ fitting from in-house scripts.
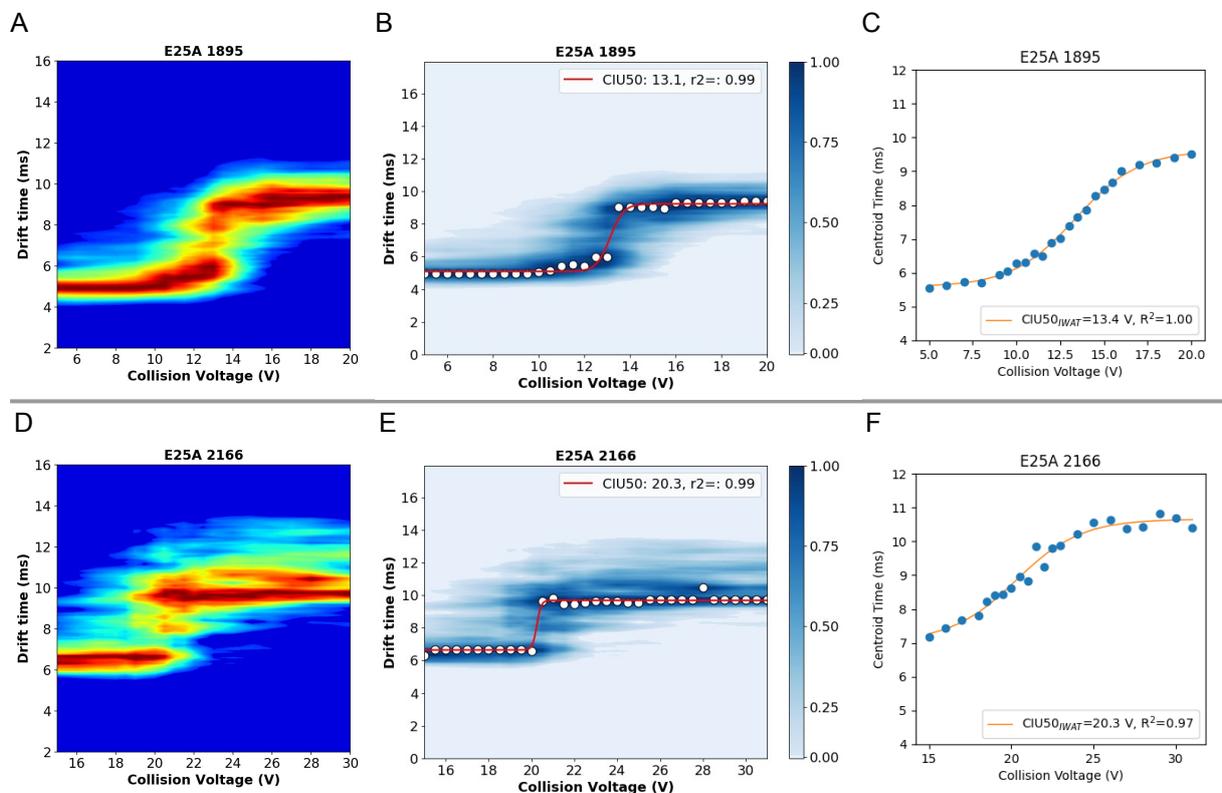
Figure S22 | CIU data for **E25A** mutant of PksACP4 for the 8+ (A, B and C) and 7+ (D, E and F) charge states. A and D) CIU plot for the 8+ charge state. B and E) CIU50$_{FD}$ fitting from CIUSuite2. C and F) CIU50$_{IWAT}$ fitting from in-house scripts.
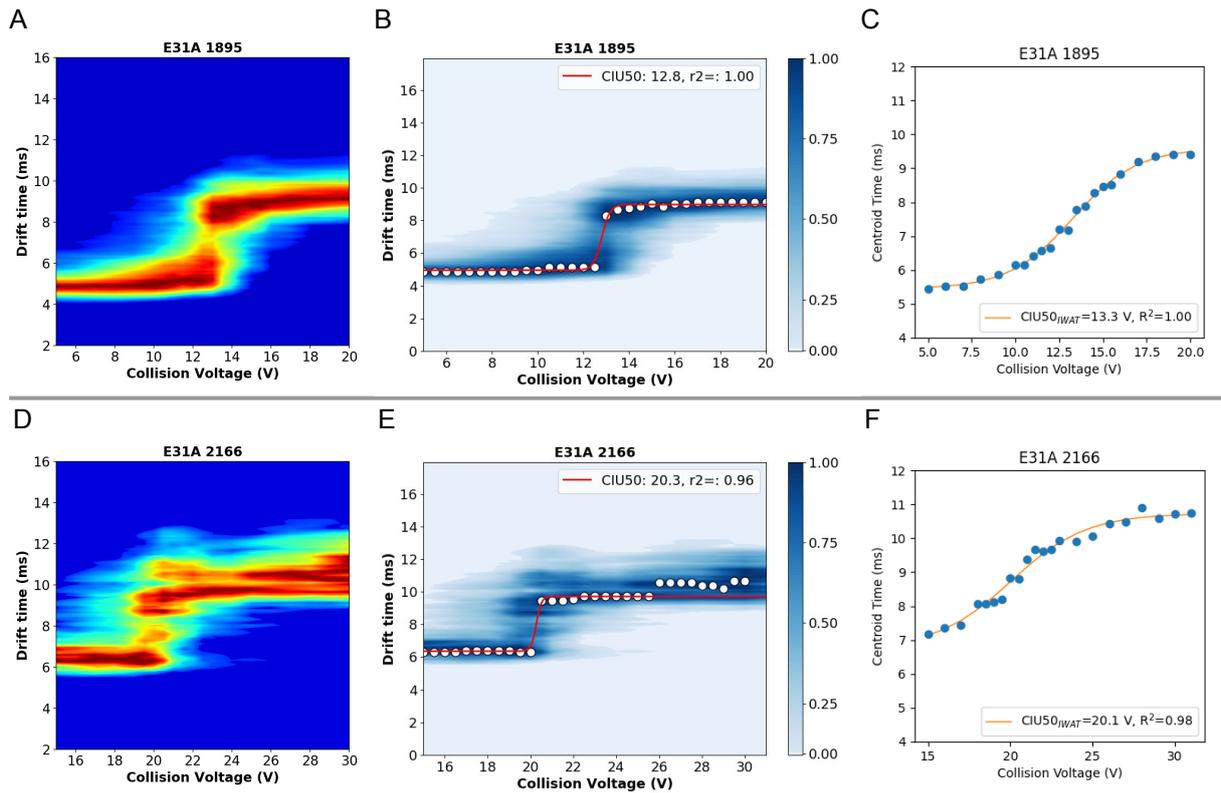
Figure S23 | CIU data for **E31A** mutant of PksACP4 for the 8+ (A, B and C) and 7+ (D, E and F) charge states. A and D) CIU plot for the 8+ charge state. B and E) CIU50$_{FD}$ fitting from CIUSuite2. C and F) CIU50$_{IWAT}$ fitting from in-house scripts.
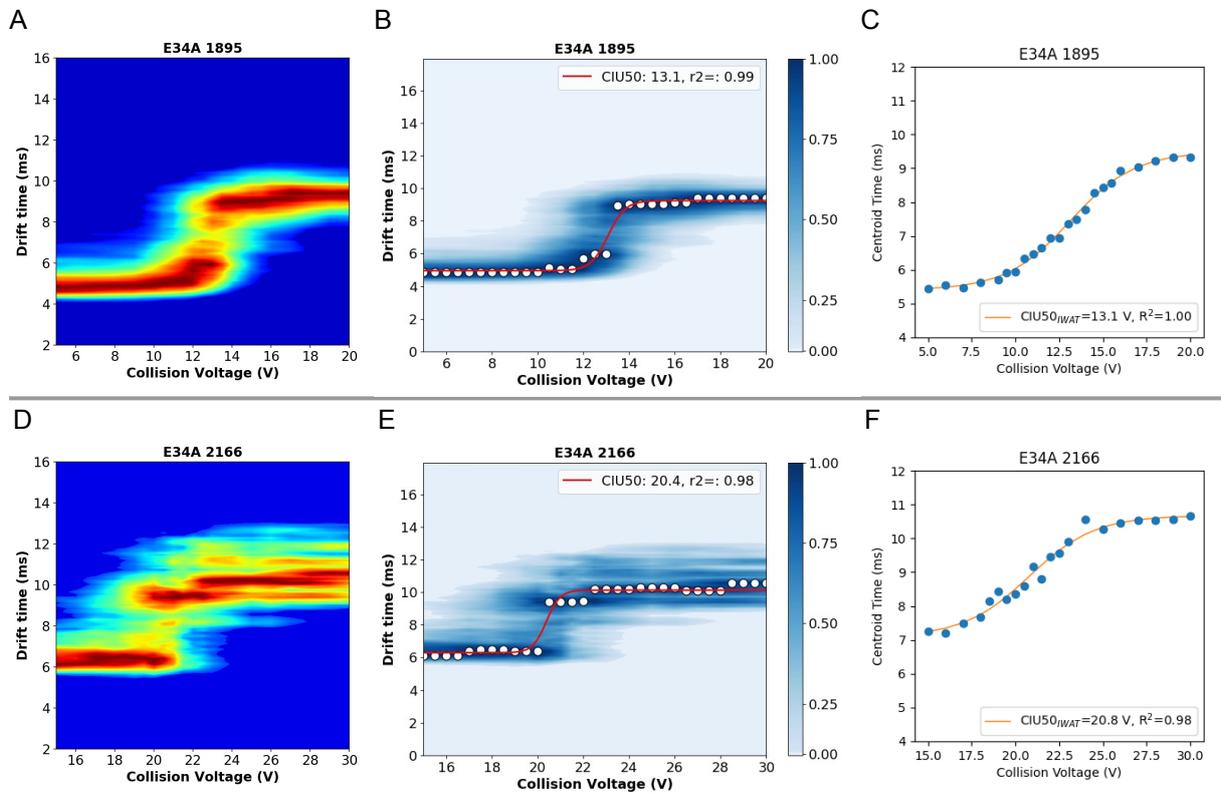
Figure S24 | CIU data for **E34A** mutant of PksACP4 for the 8+ (A, B and C) and 7+ (D, E and F) charge states. A and D) CIU plot for the 8+ charge state. B and E) CIU50$_{FD}$ fitting from CIUSuite2. C and F) CIU50$_{IWAT}$ fitting from in-house scripts.
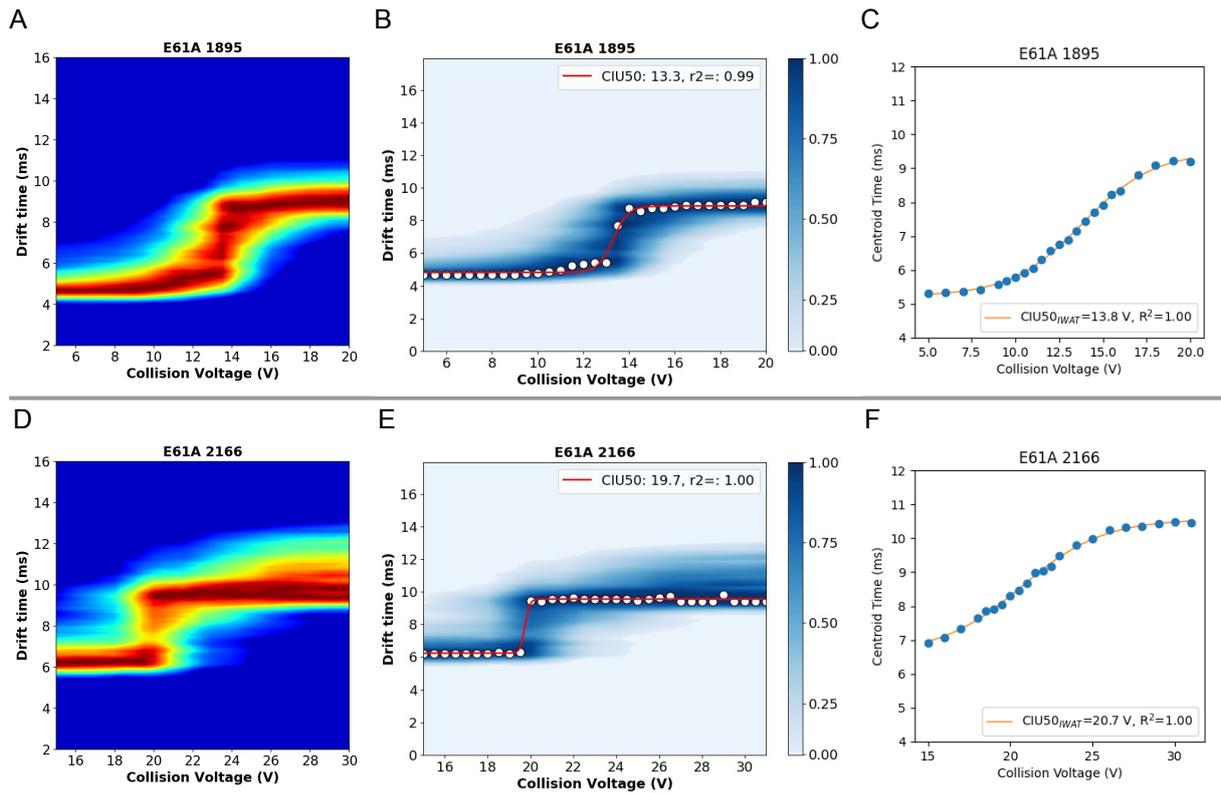
Figure S25 | CIU data for **E61A** mutant of PksACP4 for the 8+ (A, B and C) and 7+ (D, E and F) charge states. A and D) CIU plot for the 8+ charge state. B and E) CIU50$_{FD}$ fitting from CIUSuite2. C and F) CIU50$_{IWAT}$ fitting from in-house scripts.
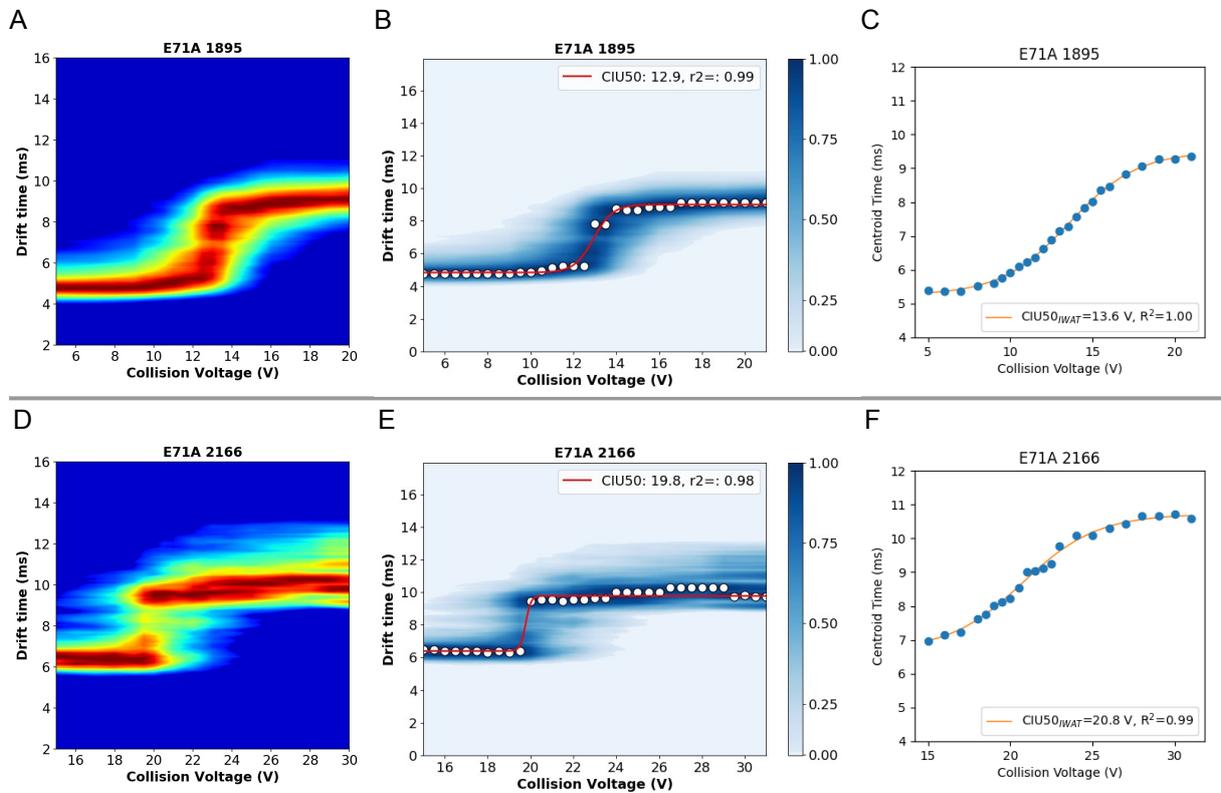
Figure S26 | CIU data for **E71A** mutant of PksACP4 for the 8+ (A, B and C) and 7+ (D, E and F) charge states. A and D) CIU plot for the 8+ charge state. B and E) CIU50$_{FD}$ fitting from CIUSuite2. C and F) CIU50$_{IWAT}$ fitting from in-house scripts.
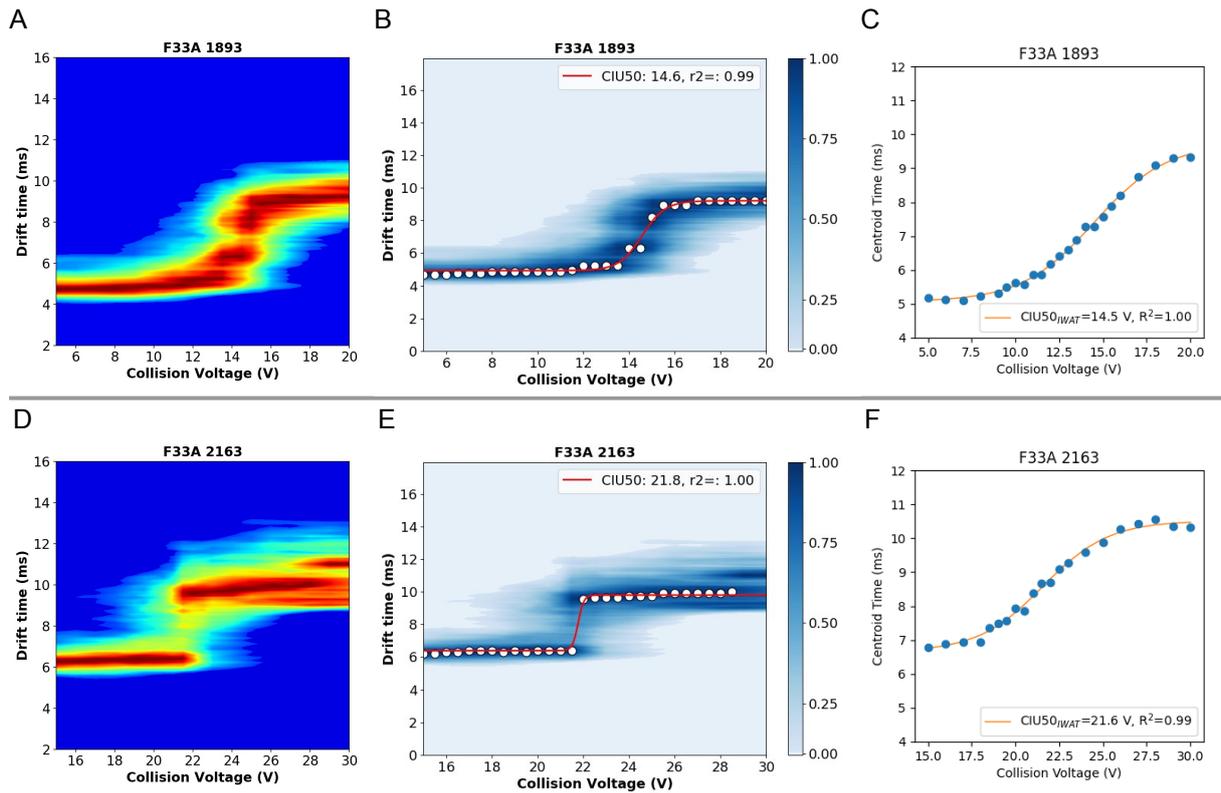
Figure S27 | CIU data for **F33A** mutant of PksACP4 for the 8+ (A, B and C) and 7+ (D, E and F) charge states. A and D) CIU plot for the 8+ charge state. B and E) CIU50$_{FD}$ fitting from CIUSuite2. C and F) CIU50$_{IWAT}$ fitting from in-house scripts.
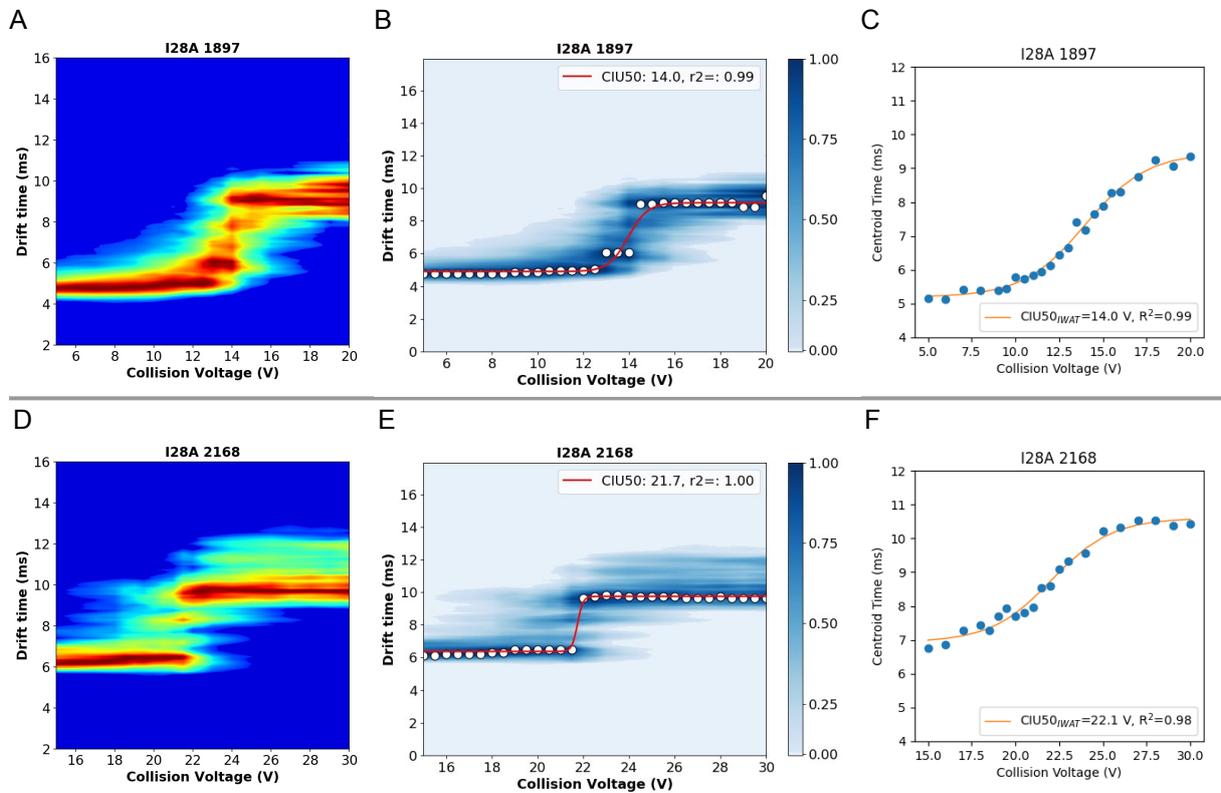
Figure S28 | CIU data for **I28A** mutant of PksACP4 for the 8+ (A, B and C) and 7+ (D, E and F) charge states. A and D) CIU plot for the 8+ charge state. B and E) CIU50$_{FD}$ fitting from CIUSuite2. C and F) CIU50$_{IWAT}$ fitting from in-house scripts.
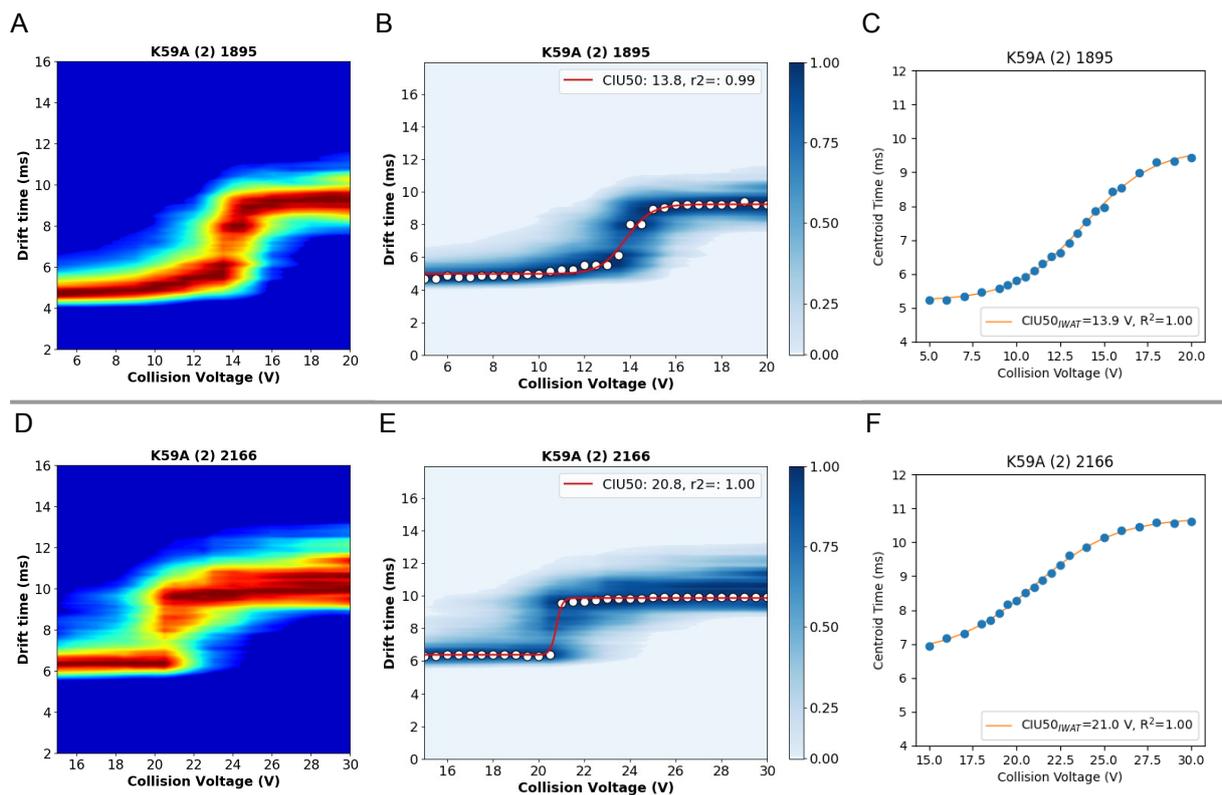
Figure S29 | CIU data for **K59A** mutant replicate **2** of PksACP4 for the 8+ (A, B and C) and 7+ (D, E and F) charge states. A and D) CIU plot for the 8+ charge state. B and E) CIU50$_{FD}$ fitting from CIUSuite2. C and F) CIU50$_{IWAT}$ fitting from in-house scripts.
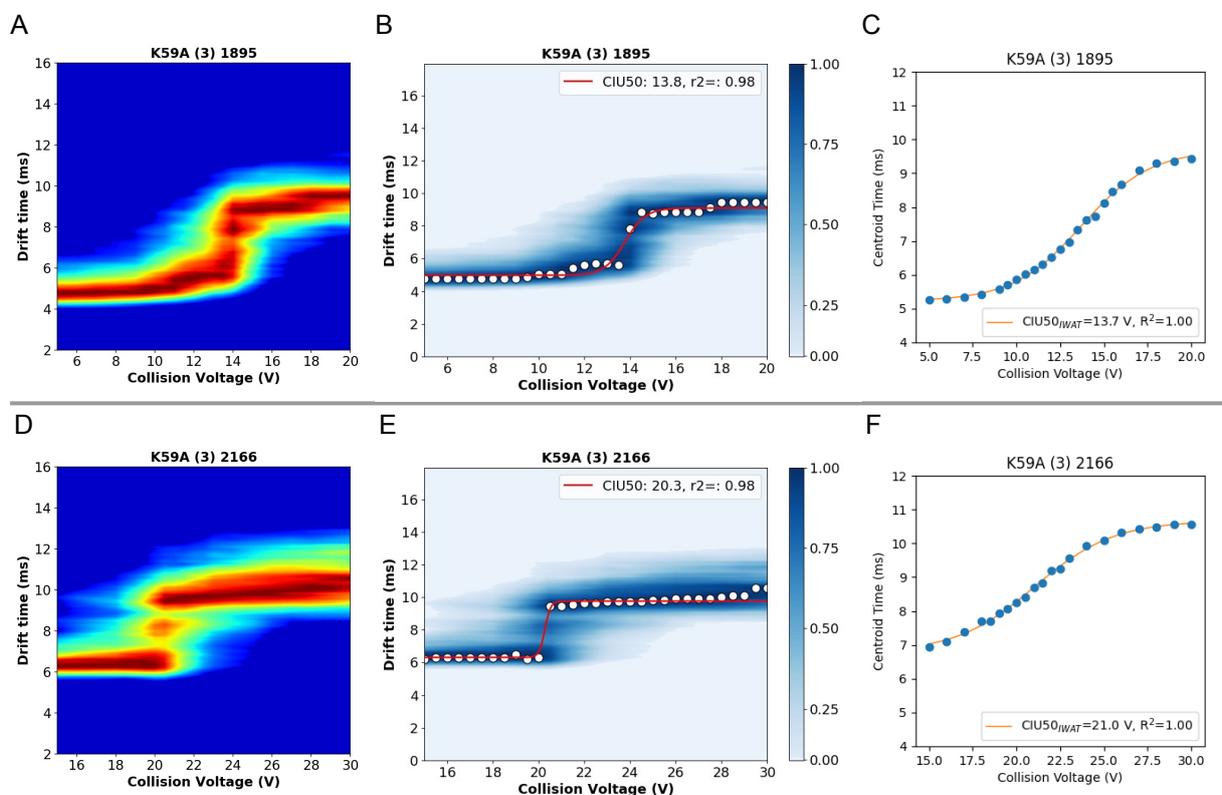
Figure S30 | CIU data for **K59A** mutant replicate **3** of PksACP4 for the 8+ (A, B and C) and 7+ (D, E and F) charge states. A and D) CIU plot for the 8+ charge state. B and E) CIU50$_{FD}$ fitting from CIUSuite2. C and F) CIU50$_{IWAT}$ fitting from in-house scripts.
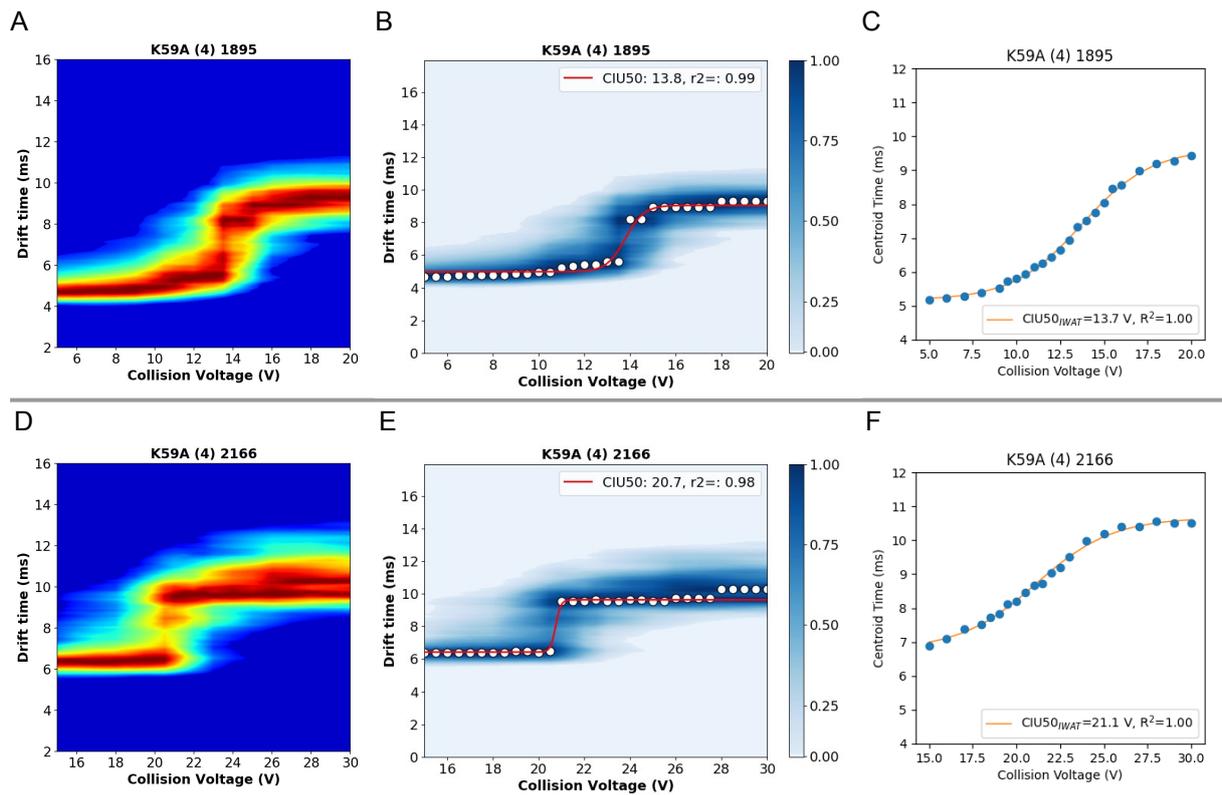
Figure S31 | CIU data for **K59A** mutant replicate **4** of PksACP4 for the 8+ (A, B and C) and 7+ (D, E and F) charge states. A and D) CIU plot for the 8+ charge state. B and E) CIU50$_{FD}$ fitting from CIUSuite2. C and F) CIU50$_{IWAT}$ fitting from in-house scripts.
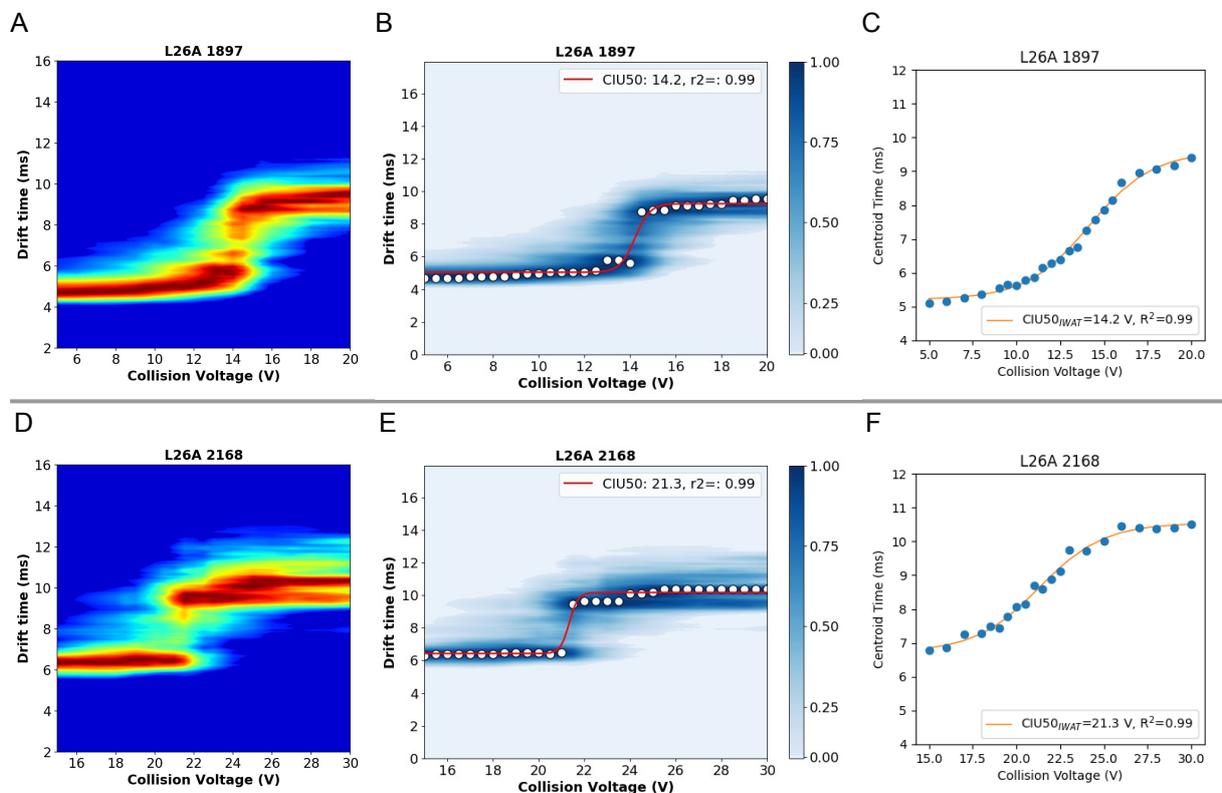
Figure S32 | CIU data for **L26A** mutant of PksACP4 for the 8+ (A, B and C) and 7+ (D, E and F) charge states. A and D) CIU plot for the 8+ charge state. B and E) CIU50$_{FD}$ fitting from CIUSuite2. C and F) CIU50$_{IWAT}$ fitting from in-house scripts.
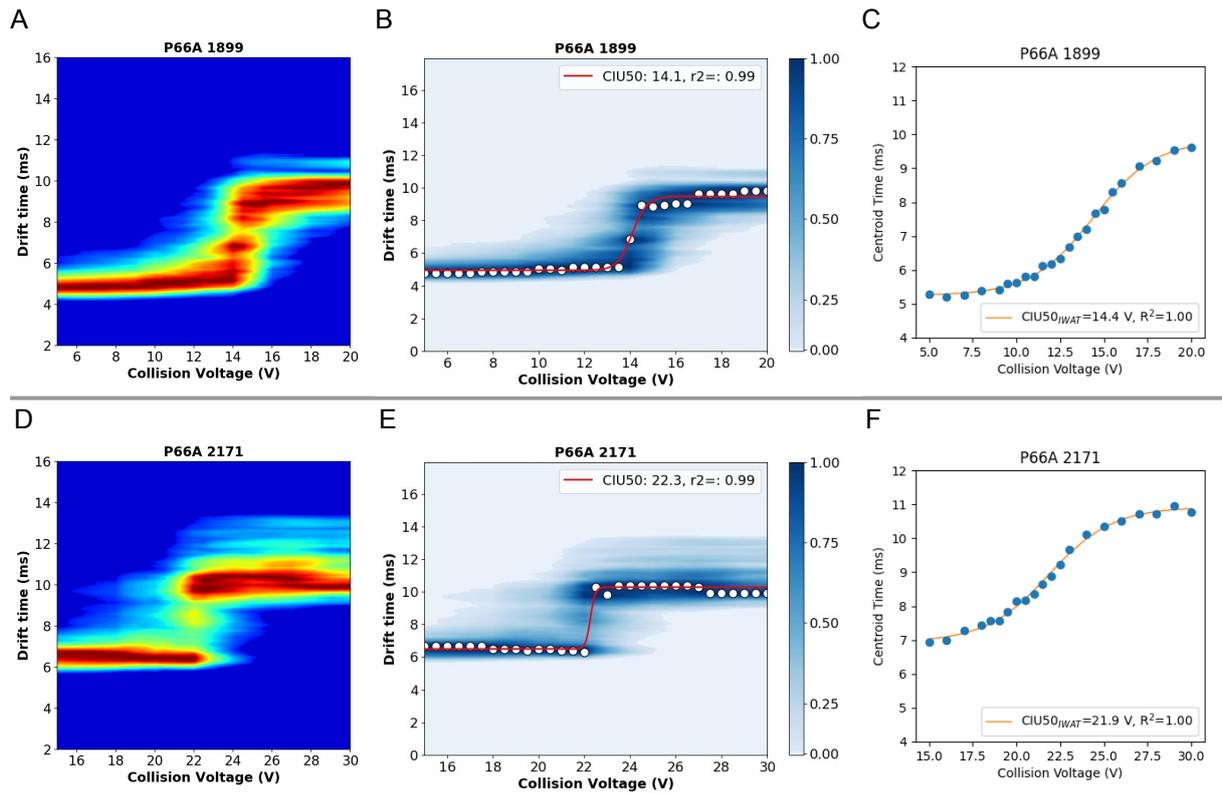
Figure S33 | CIU data for **P66A** mutant of PksACP4 for the 8+ (A, B and C) and 7+ (D, E and F) charge states.
A and D) CIU plot for the 8+ charge state. B and E) CIU50$_{FD}$ fitting from CIUSuite2. C and F) CIU50$_{IWAT}$ fitting from in-house scripts.
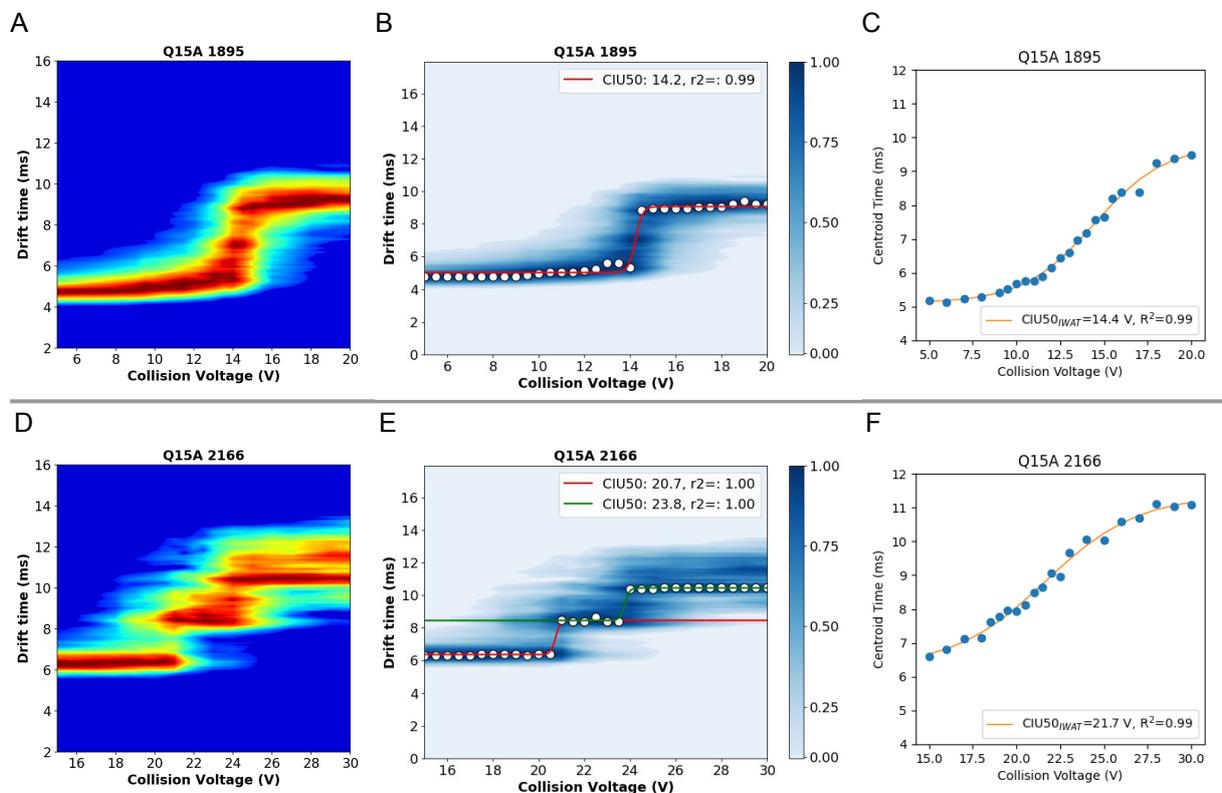
Figure S34 | CIU data for **Q15A** mutant of PksACP4 for the 8+ (A, B and C) and 7+ (D, E and F) charge states.
A and D) CIU plot for the 8+ charge state. B and E) CIU50$_{FD}$ fitting from CIUSuite2. C and F) CIU50$_{IWAT}$ fitting from in-house scripts.
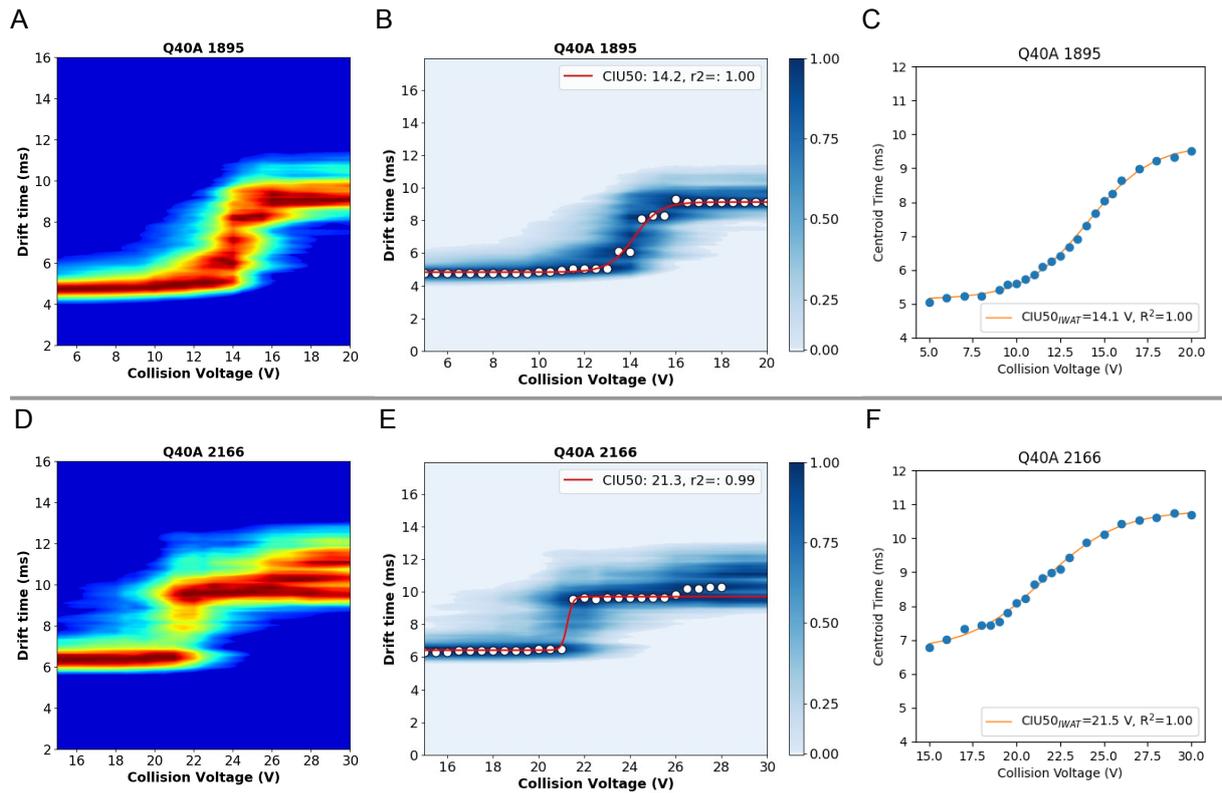
Figure S35 | CIU data for **Q40A** mutant of PksACP4 for the 8+ (A, B and C) and 7+ (D, E and F) charge states. A and D) CIU plot for the 8+ charge state. B and E) CIU50$_{FD}$ fitting from CIUSuite2. C and F) CIU50$_{IWAT}$ fitting from in-house scripts.
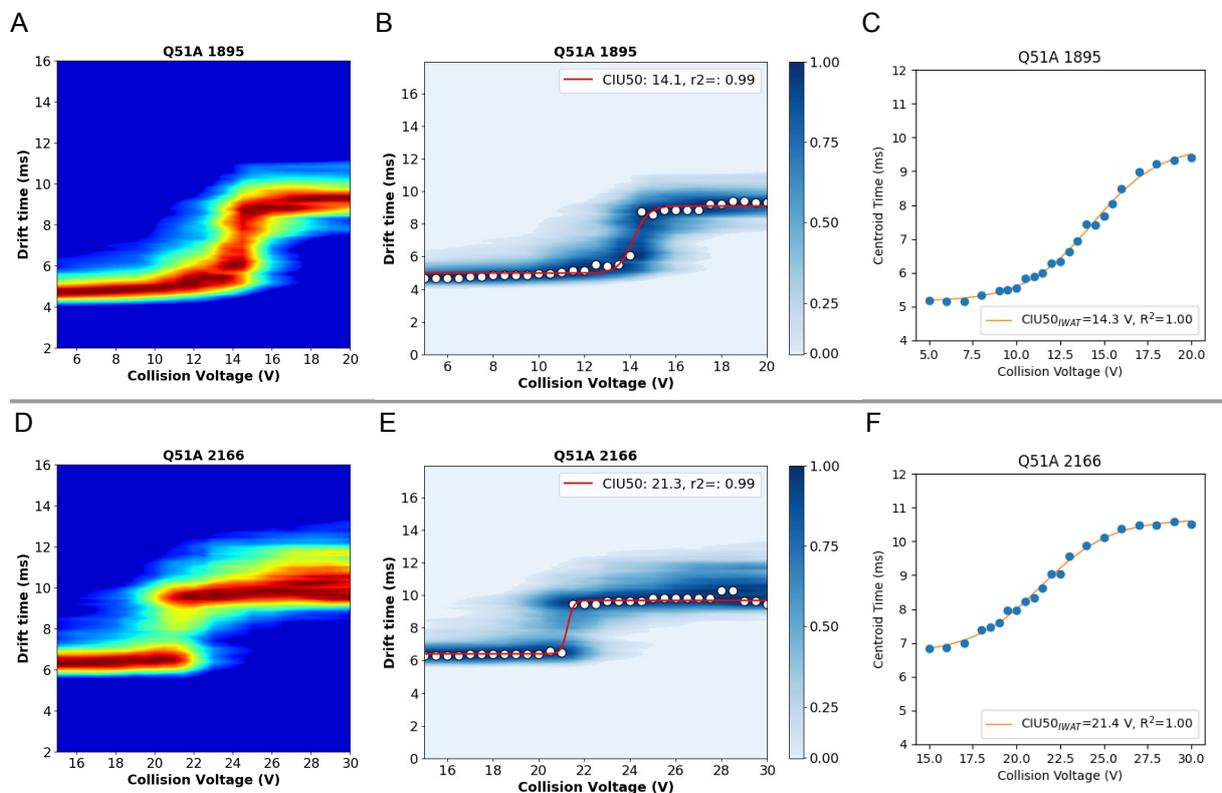
Figure S36 | CIU data for **Q51A** mutant of PksACP4 for the 8+ (A, B and C) and 7+ (D, E and F) charge states. A and D) CIU plot for the 8+ charge state. B and E) CIU50$_{FD}$ fitting from CIUSuite2. C and F) CIU50$_{IWAT}$ fitting from in-house scripts.
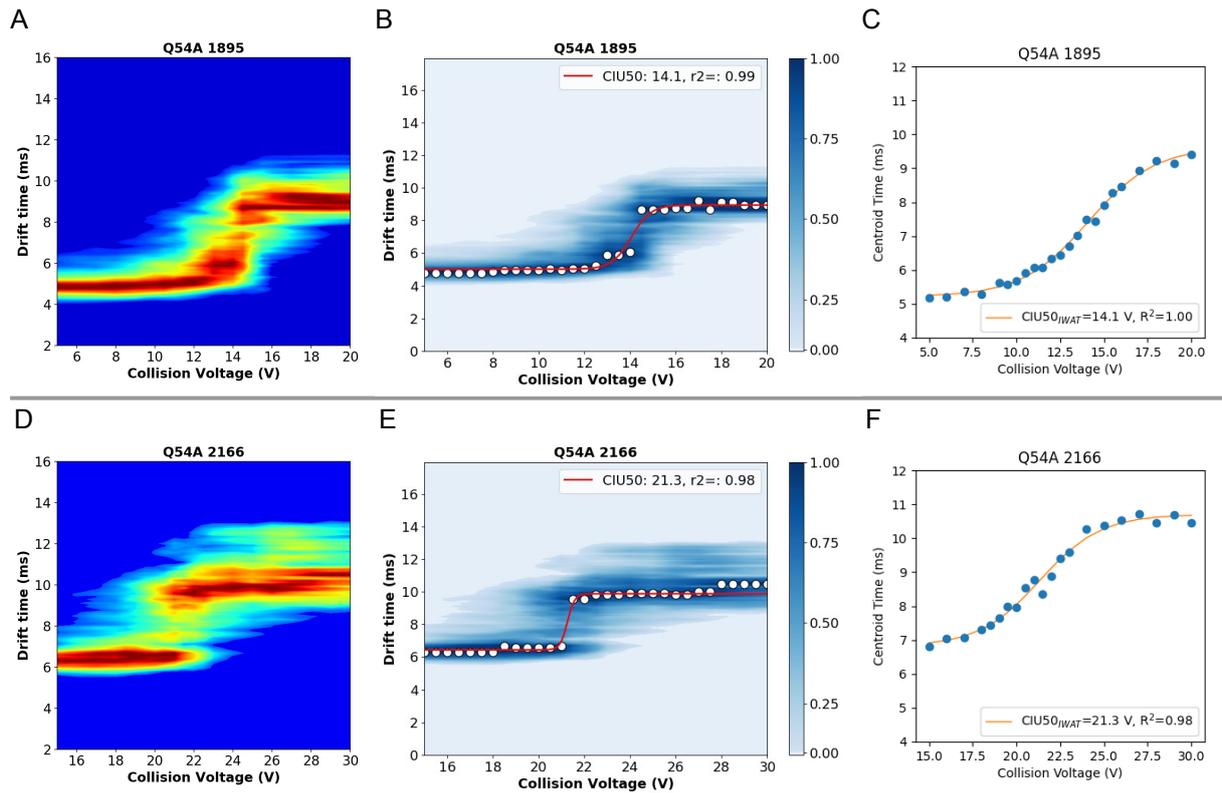
Figure S37 | CIU data for **Q54A** mutant of PksACP4 for the 8+ (A, B and C) and 7+ (D, E and F) charge states. A and D) CIU plot for the 8+ charge state. B and E) CIU50$_{FD}$ fitting from CIUSuite2. C and F) CIU50$_{IWAT}$ fitting from in-house scripts.
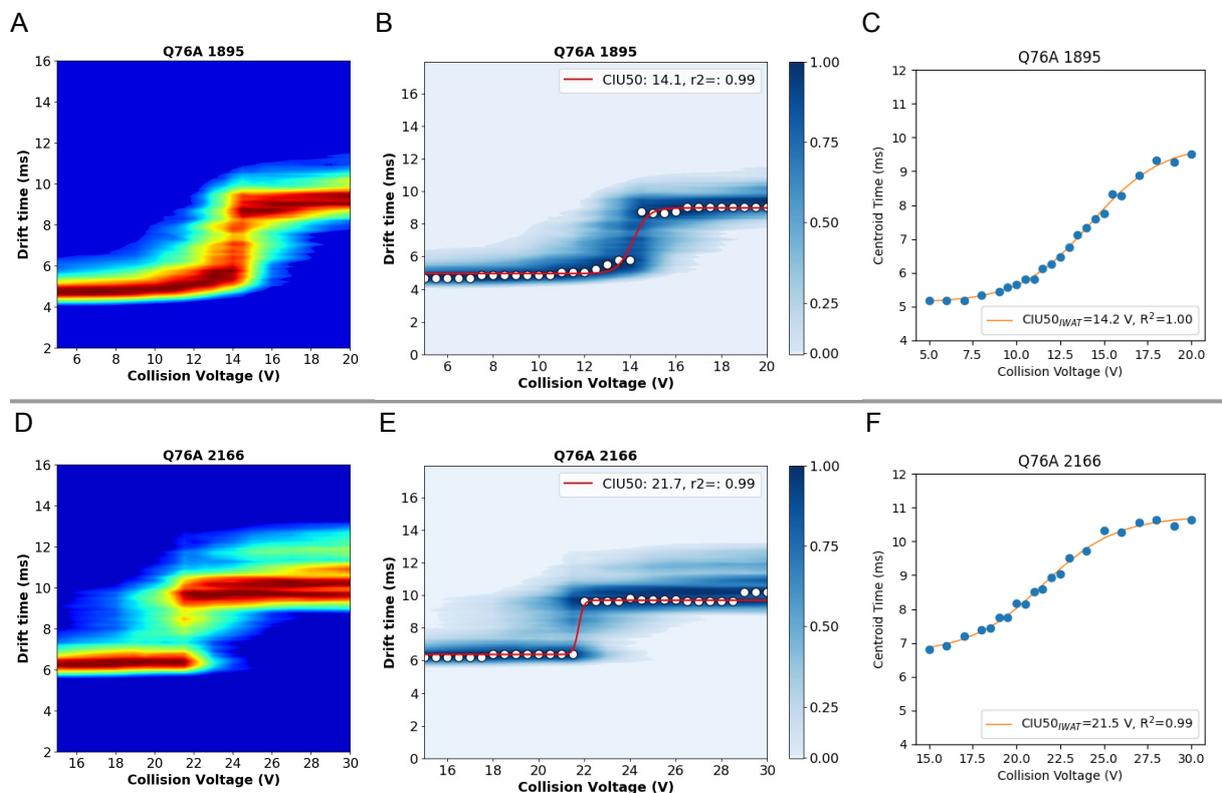
Figure S38 | CIU data for **Q76A** mutant of PksACP4 for the 8+ (A, B and C) and 7+ (D, E and F) charge states. A and D) CIU plot for the 8+ charge state. B and E) CIU50$_{FD}$ fitting from CIUSuite2. C and F) CIU50$_{IWAT}$ fitting from in-house scripts.
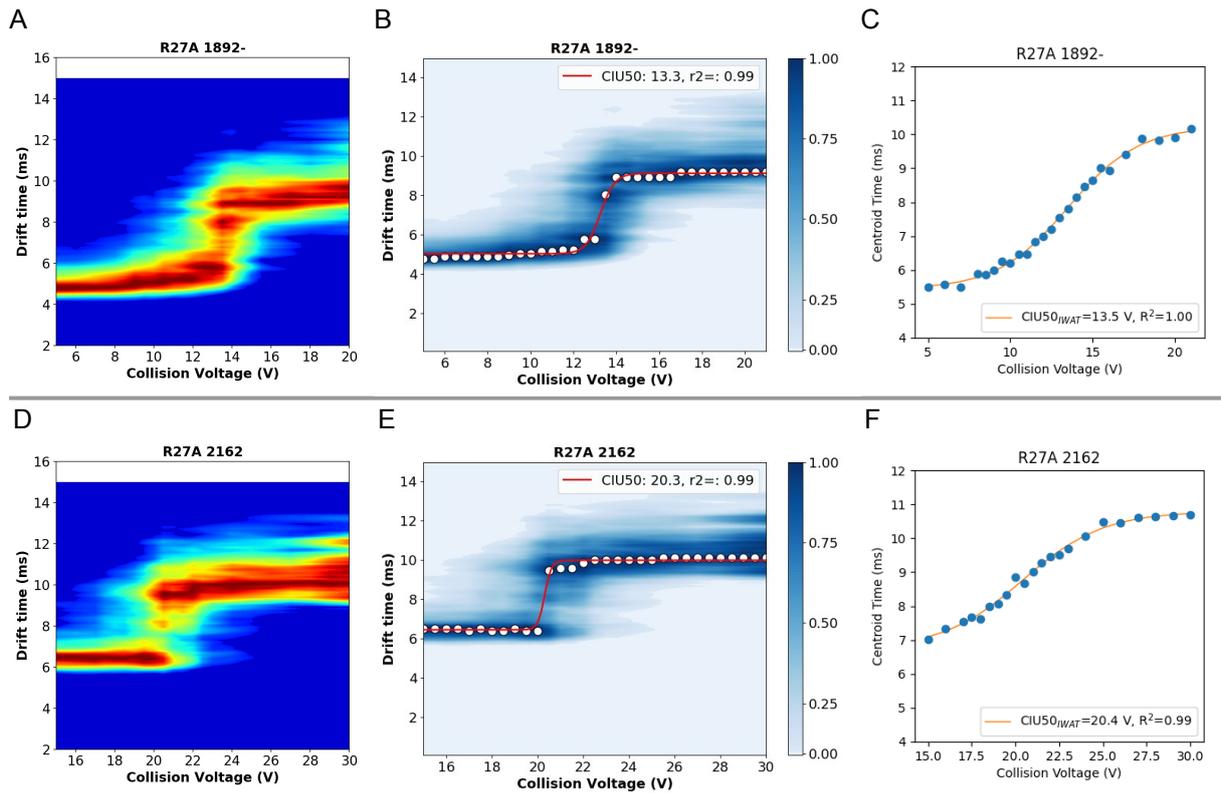
Figure S39 | CIU data for **R27A** mutant of PksACP4 for the 8+ (A, B and C) and 7+ (D, E and F) charge states. A and D) CIU plot for the 8+ charge state. B and E) CIU50$_{FD}$ fitting from CIUSuite2. C and F) CIU50$_{IWAT}$ fitting from in-house scripts.
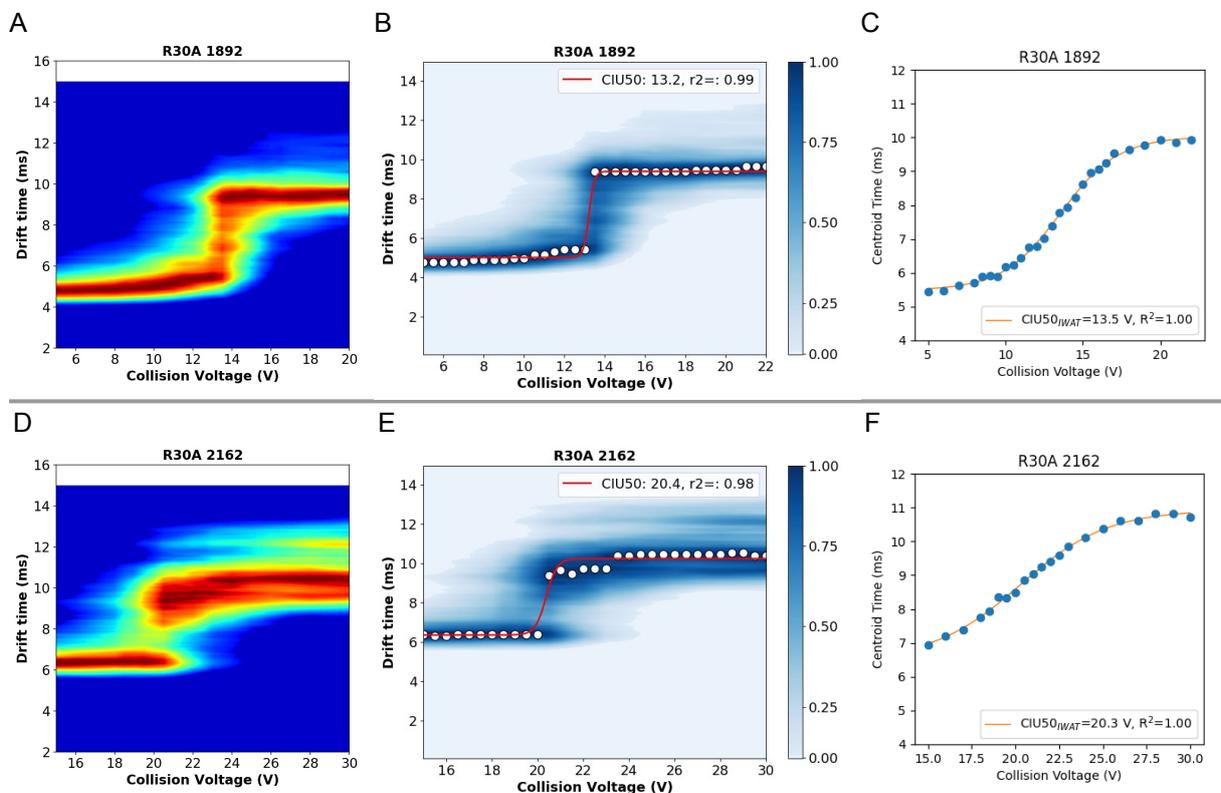
Figure S40 | CIU data for **R30A** mutant of PksACP4 for the 8+ (A, B and C) and 7+ (D, E and F) charge states. A and D) CIU plot for the 8+ charge state. B and E) CIU50$_{FD}$ fitting from CIUSuite2. C and F) CIU50$_{IWAT}$ fitting from in-house scripts.
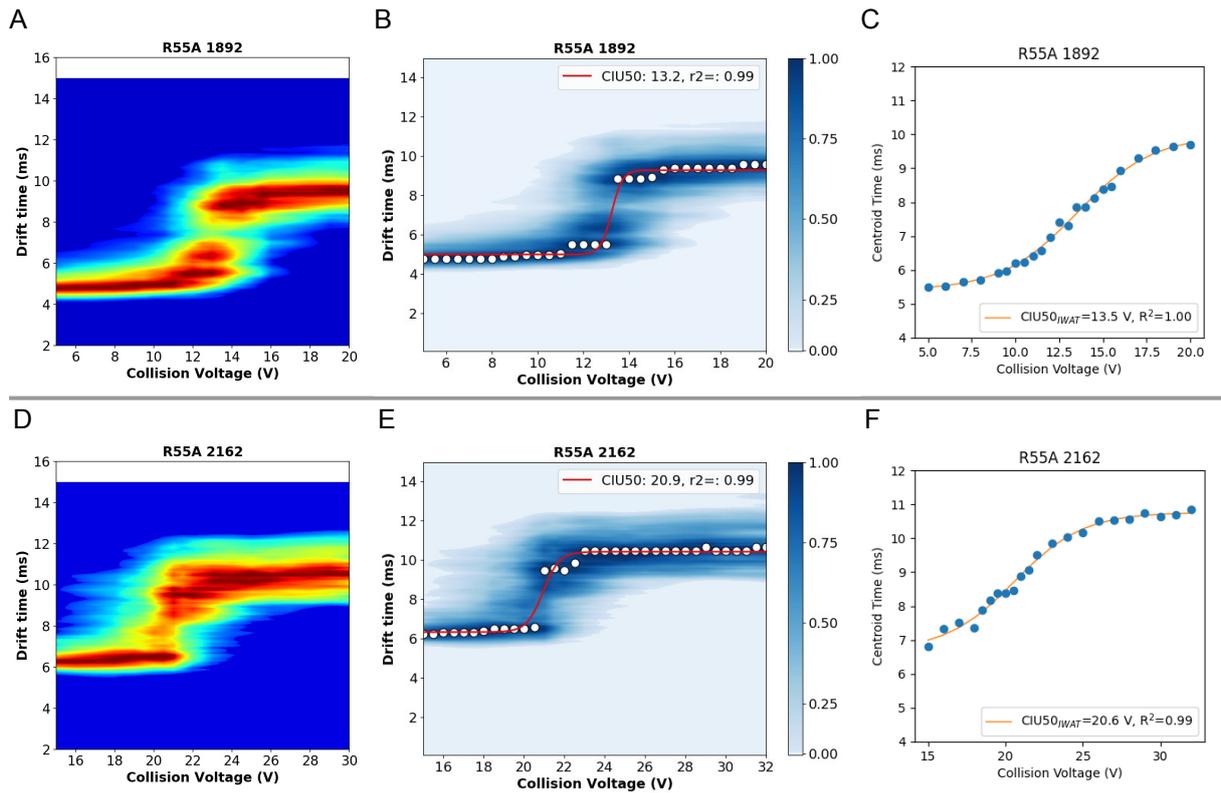
Figure S41 | CIU data for **R55A** mutant of PksACP4 for the 8+ (A, B and C) and 7+ (D, E and F) charge states. A and D) CIU plot for the 8+ charge state. B and E) CIU50$_{FD}$ fitting from CIUSuite2. C and F) CIU50$_{IWAT}$ fitting from in-house scripts.
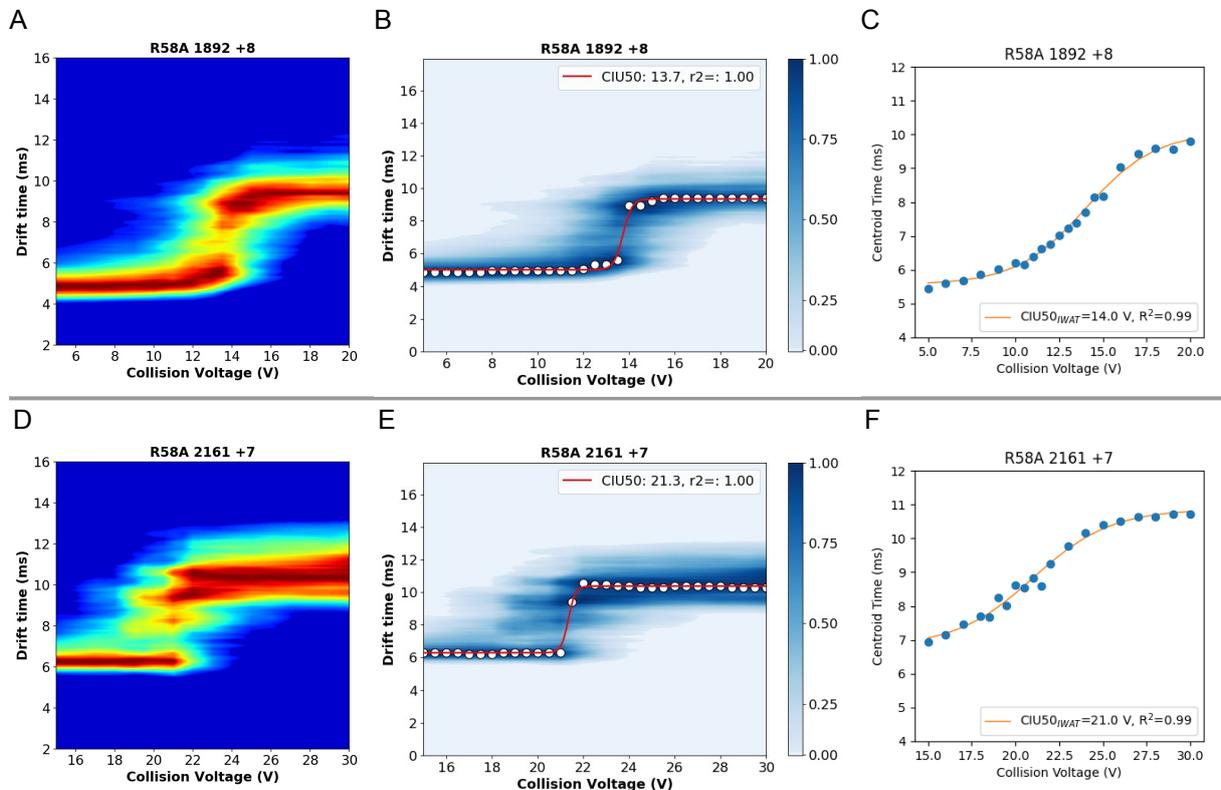
Figure S42 | CIU data for **R58A** mutant of PksACP4 for the 8+ (A, B and C) and 7+ (D, E and F) charge states. A and D) CIU plot for the 8+ charge state. B and E) CIU50$_{FD}$ fitting from CIUSuite2. C and F) CIU50$_{IWAT}$ fitting from in-house scripts.
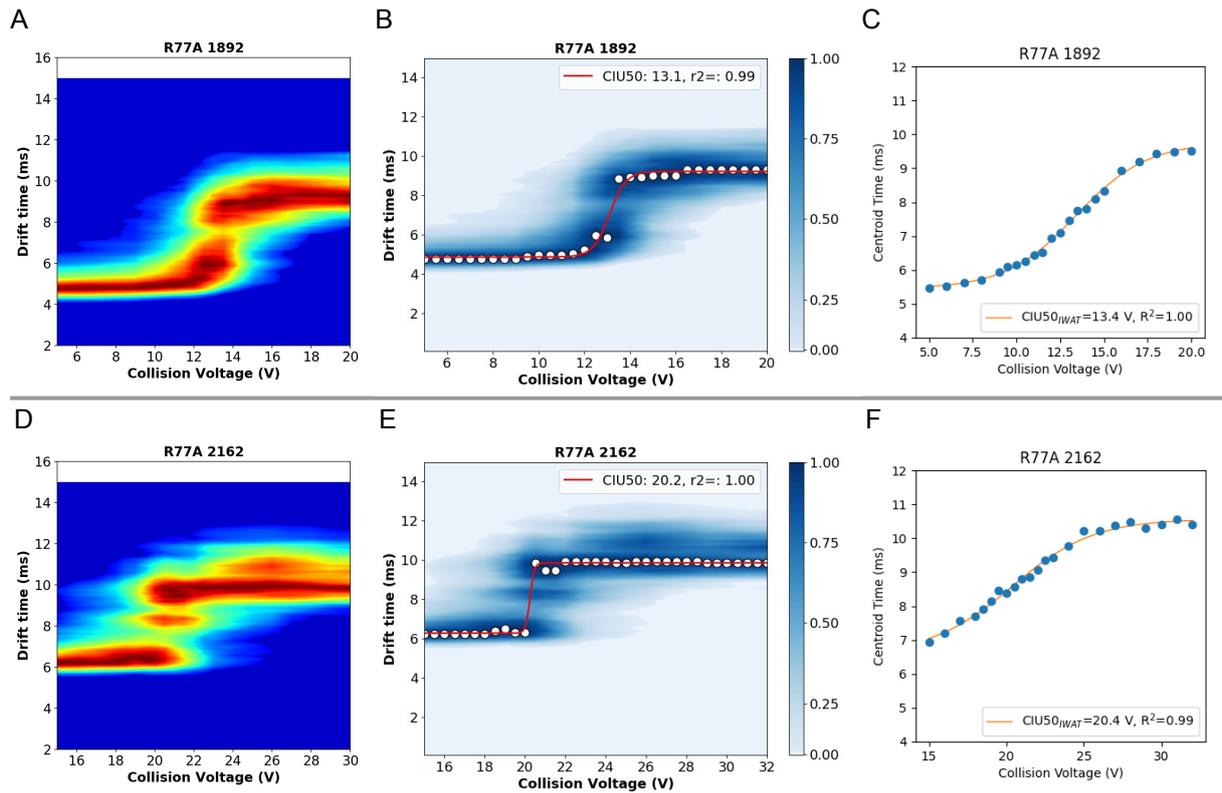
Figure S43 | CIU data for **R77A** mutant of PksACP4 for the 8+ (A, B and C) and 7+ (D, E and F) charge states. A and D) CIU plot for the 8+ charge state. B and E) CIU50$_{FD}$ fitting from CIUSuite2. C and F) CIU50$_{IWAT}$ fitting from in-house scripts.
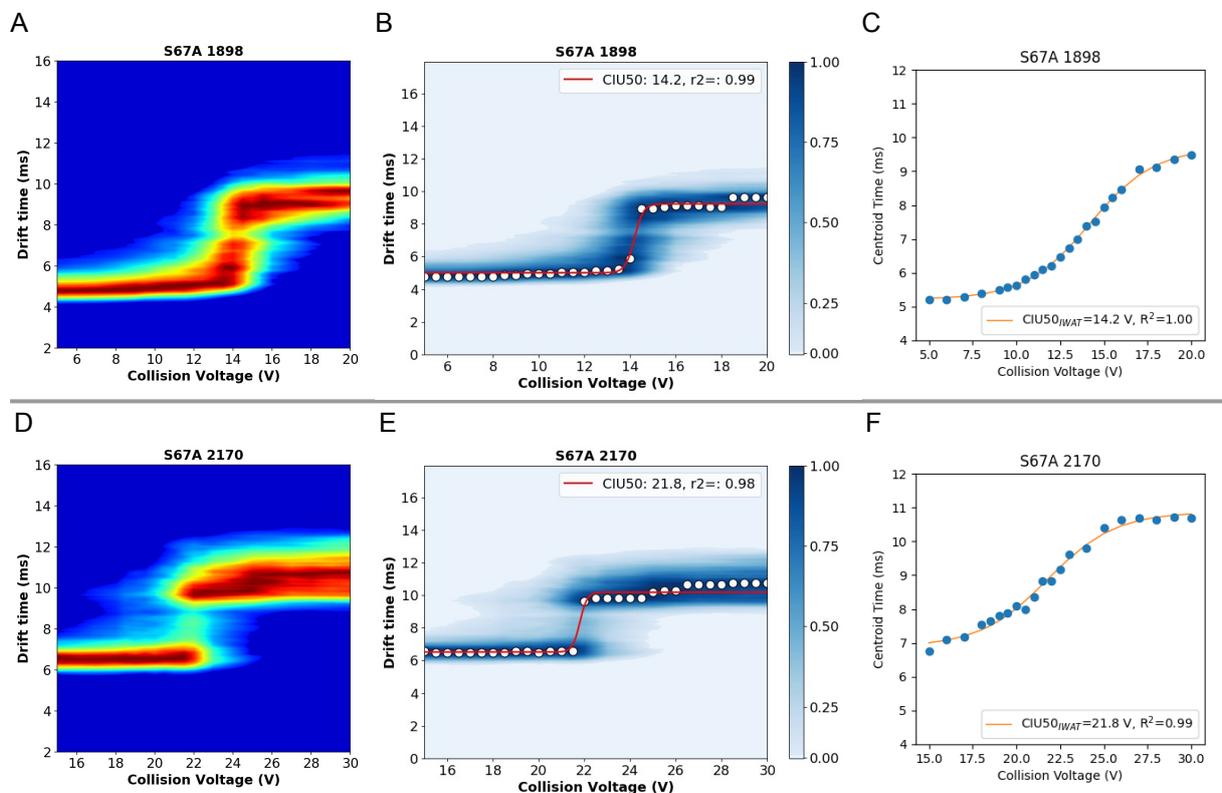
Figure S44 | CIU data for **S67A** mutant of PksACP4 for the 8+ (A, B and C) and 7+ (D, E and F) charge states. A and D) CIU plot for the 8+ charge state. B and E) CIU50$_{FD}$ fitting from CIUSuite2. C and F) CIU50$_{IWAT}$ fitting from in-house scripts.
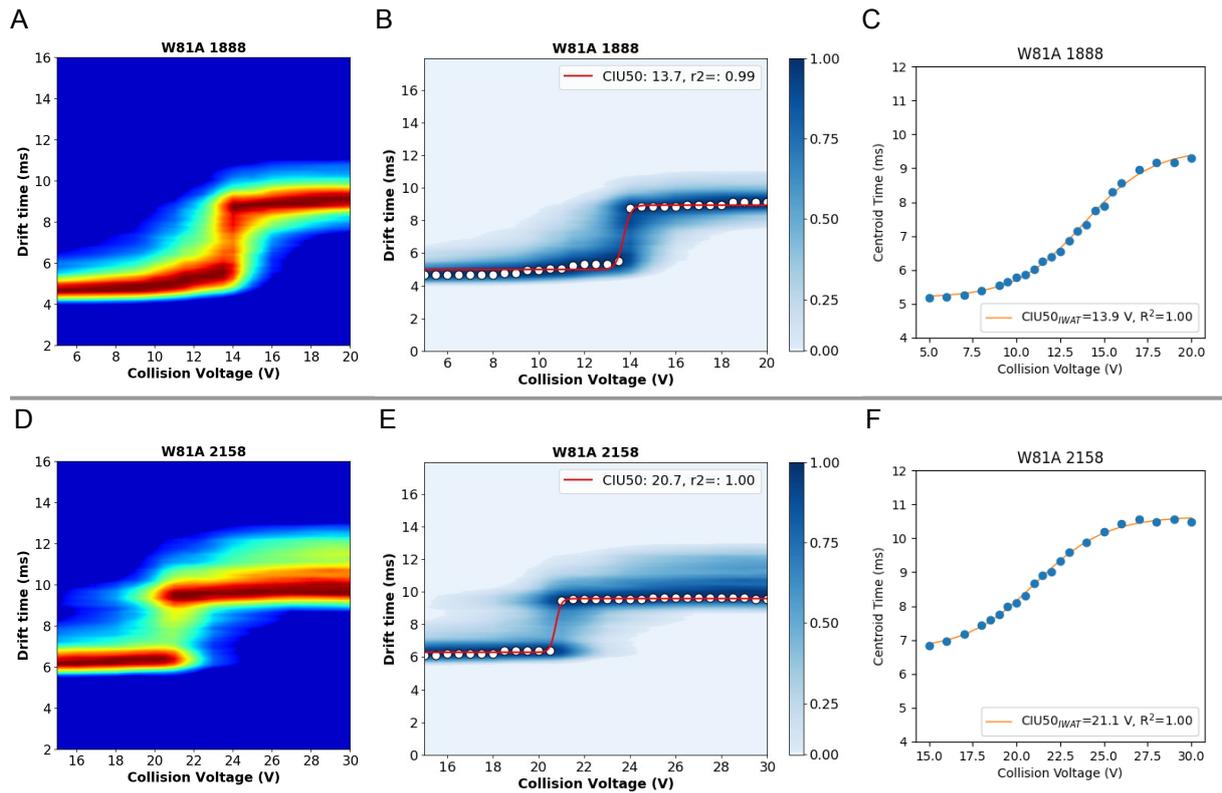
Figure S45 | CIU data for **W81A** mutant of PksACP4 for the 8+ (A, B and C) and 7+ (D, E and F) charge states. A and D) CIU plot for the 8+ charge state. B and E) CIU50$_{FD}$ fitting from CIUSuite2. C and F) CIU50$_{IWAT}$ fitting from in-house scripts.
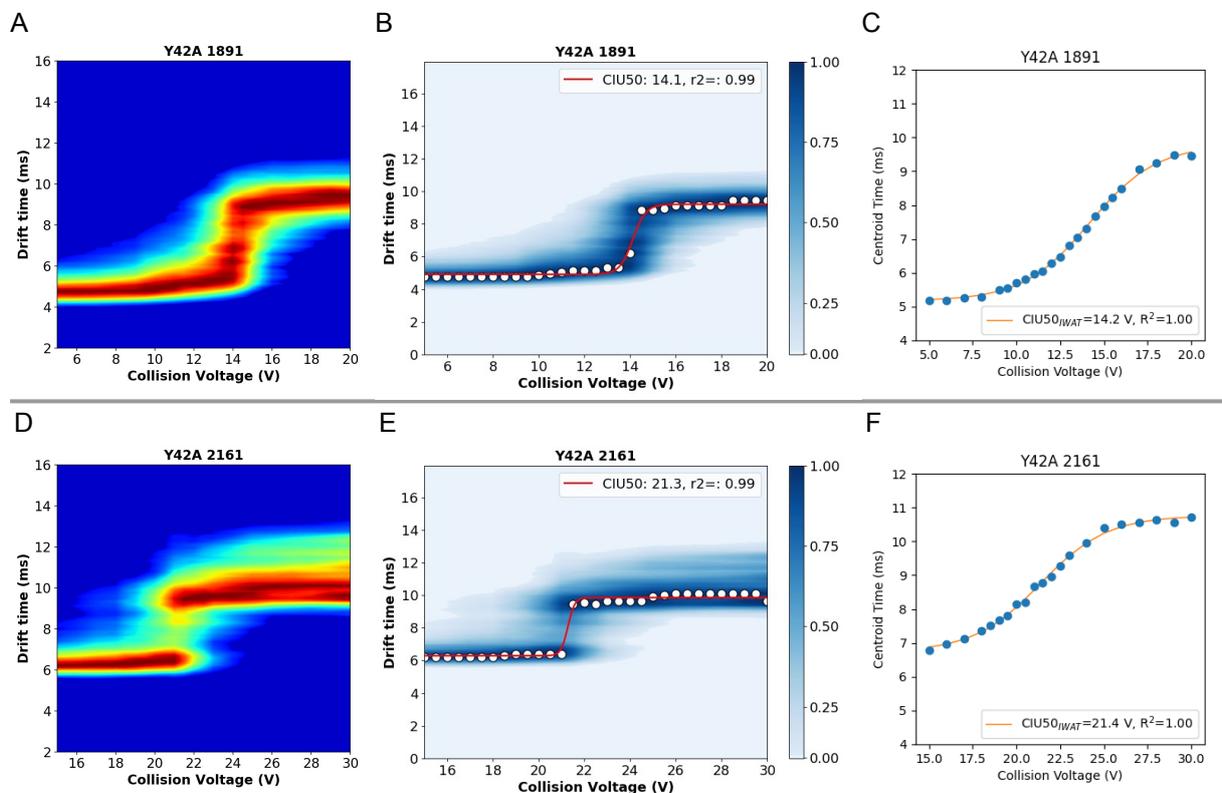
Figure S46 | CIU data for **Y42A** mutant of PksACP4 for the 8+ (A, B and C) and 7+ (D, E and F) charge states. A and D) CIU plot for the 8+ charge state. B and E) CIU50$_{FD}$ fitting from CIUSuite2. C and F) CIU50$_{IWAT}$ fitting from in-house scripts.
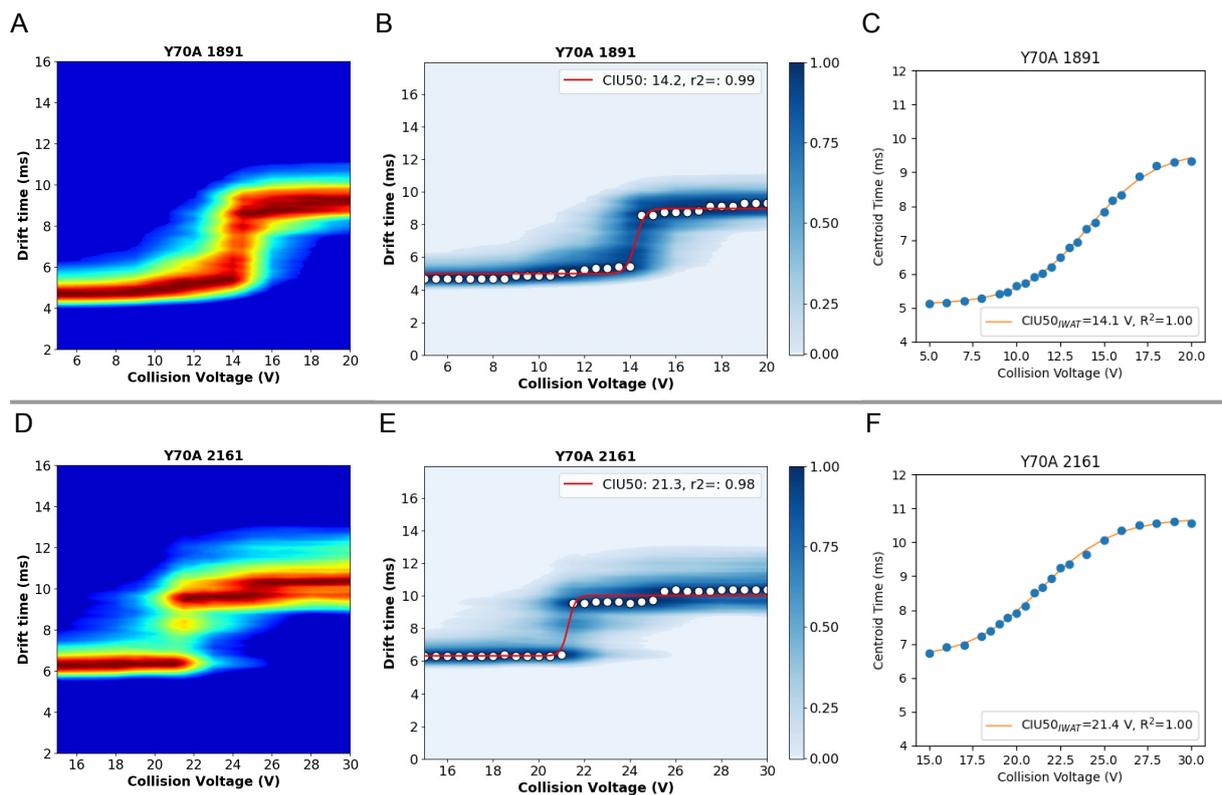
Figure S47 | CIU data for **Y70A** mutant of PksACP4 for the 8+ (A, B and C) and 7+ (D, E and F) charge states. A and D) CIU plot for the 8+ charge state. B and E) CIU50$_{FD}$ fitting from CIUSuite2. C and F) CIU50$_{IWAT}$ fitting from in-house scripts.
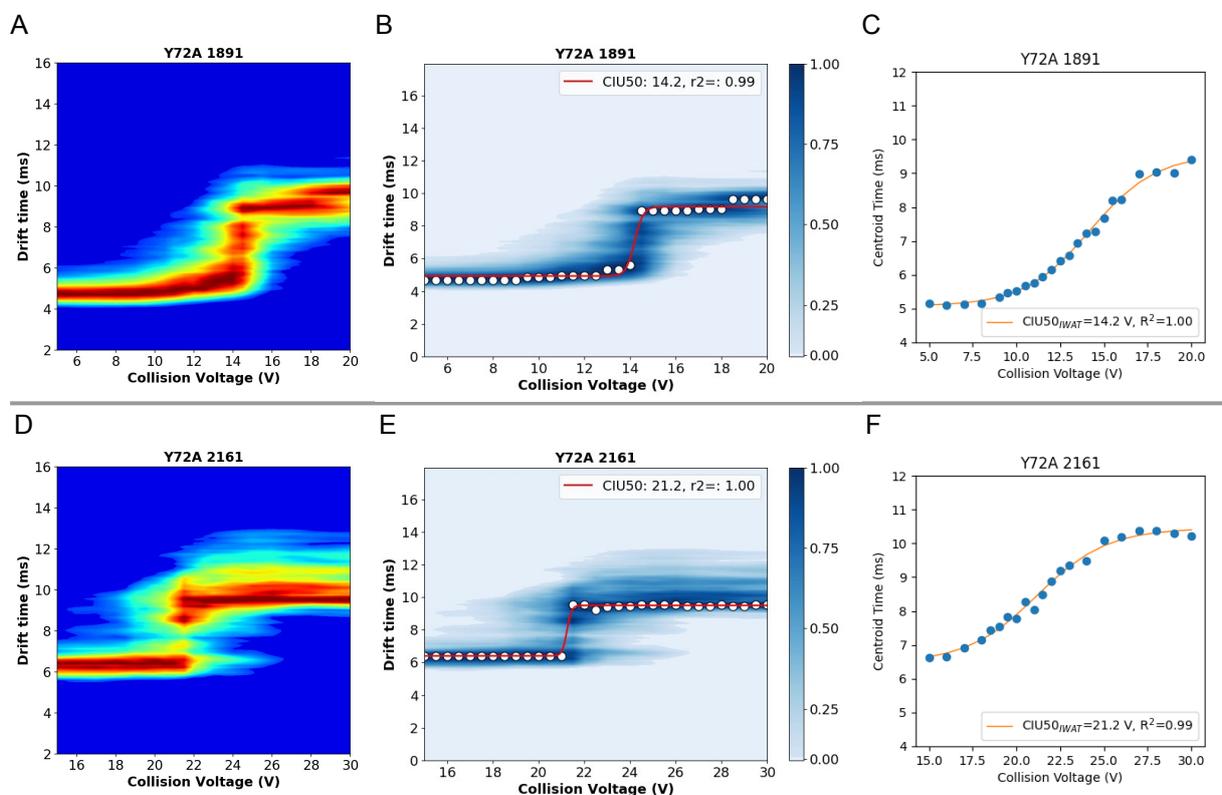
Figure S48 | CIU data for **Y72A** mutant of PksACP4 for the 8+ (A, B and C) and 7+ (D, E and F) charge states. A and D) CIU plot for the 8+ charge state. B and E) CIU50$_{FD}$ fitting from CIUSuite2. C and F) CIU50$_{IWAT}$ fitting from in-house scripts.Molecular Dynamics

Benchmarking ChargePlacer
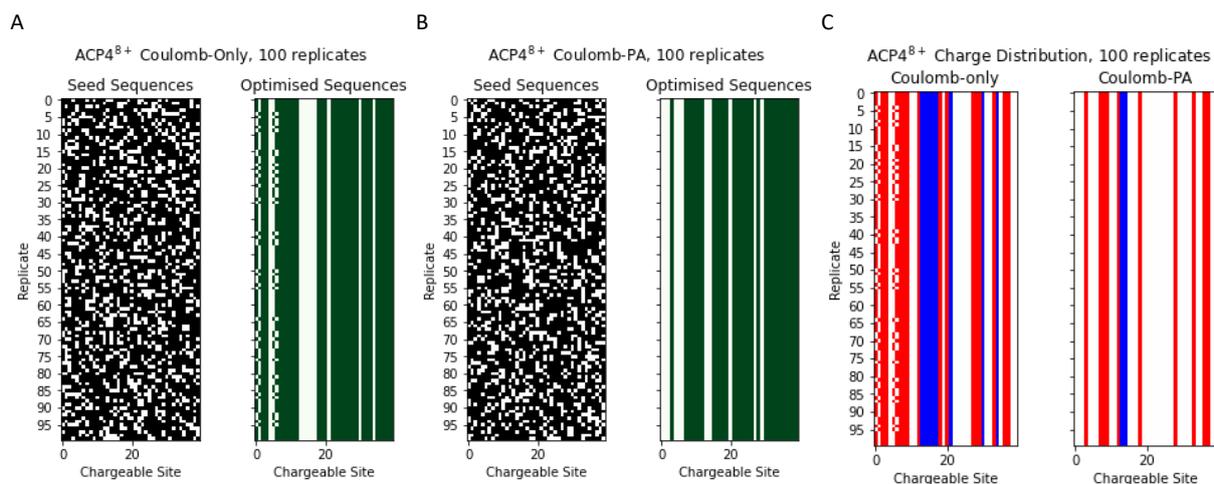
Figure S49 | Replicates (n=100) of ChargePlacer performed on ACP4 8+ charge state.
A) Coulomb-only and B) Coulomb-PA modes. A and B) Proton patterns from randomised seed sequences (black) and optimised sequences (green) show that both modes give highly reproducible convergence. C) Comparing the resulting charge patterns of these optimised sequences show that the Coulomb-only method gives higher zwitterionic character. Charges: -1, blue; 0, white; 1, red.
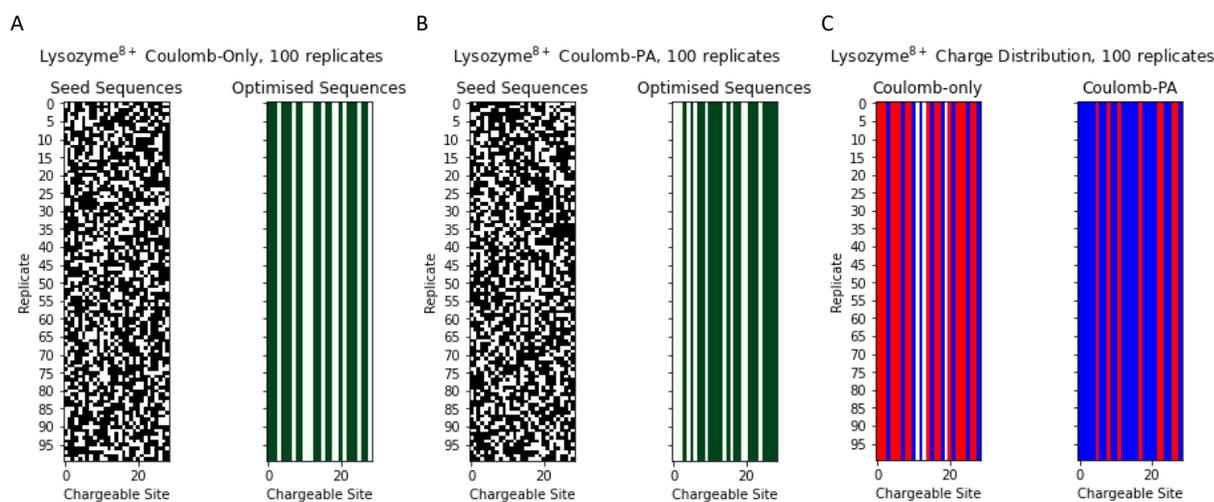
Figure S50 | Replicates (n=100) of ChargePlacer performed on Lysozyme (1AKI) 8+ charge state.
A) Coulomb-only and B) Coulomb-PA modes. A and B) Proton patterns from randomised seed sequences (black) and optimised sequences (green) show that both modes give highly reproducible convergence. C) Comparing the resulting charge patterns of these optimised sequences show that the Coulomb-PA method gives higher zwitterionic character here. Charges: -1, blue; 0, white; 1, red.
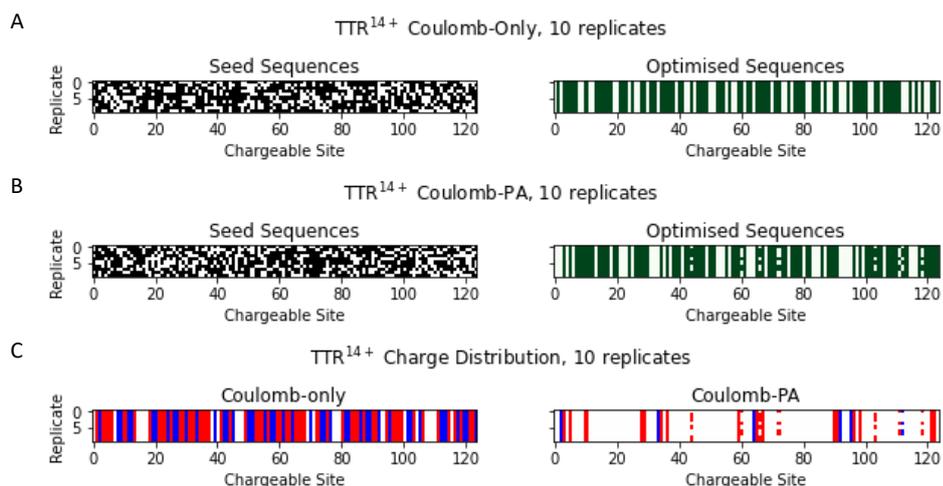
Figure S51 | Replicates (n=10) of ChargePlacer performed on TTR (3GRG) 14+ charge state.
A) Coulomb-only and B) Coulomb-PA modes. A and B) Proton patterns from randomised seed sequences (black) and optimised sequences (green) show that both modes give highly reproducible convergence. C) Comparing the resulting charge patterns of these optimised sequences show that the Coulomb-only method gives significantly higher zwitterionic character. Charges: -1, blue; 0, white; 1, red.

Table S6 | Output from ChargePlacer benchmarking with ACP4, lysozyme and TTR.
Corresponds to proton/charge patterns shown in Figures S49, S50 and S51, respectively. †Where more than one charge pattern was observed, the number of replicates with each pattern are shown in brackets. ‡The proton patterns for ACP4 are listed by residue in Table S8.

Protein	Charge	PDB	# Reps†	# Sites	# H+	Steps /shunt	Coulomb-only mode (kJ/mol)			Coulomb-PA mode (kJ/mol)		
							E_{Coul}	E_{PA}	# shunts	E_{Coul}	E_{PA}	# shunts
ACP4‡	8+	—	100 (34)	40	29	319	-1029.51	34496.0	8.8 ± 1.3	657.25	37401.0	9.8 ± 1.6
			(66)				-1035.06	34424.6				
Lysozyme	8+	1AKI	100	29	18	198	-961.27	17908.1	8.5 ± 1.5	1583.48	22517.5	8.3 ± 1.3
TTR	14+	3GRG	10 (4)	124	78	3588	-10054.71	84685.7	30.2 ± 3.1	602.19	103781.5	31.4 ± 2.6
			(6)							1070.16	104268.0	

In-silico alanine scanning
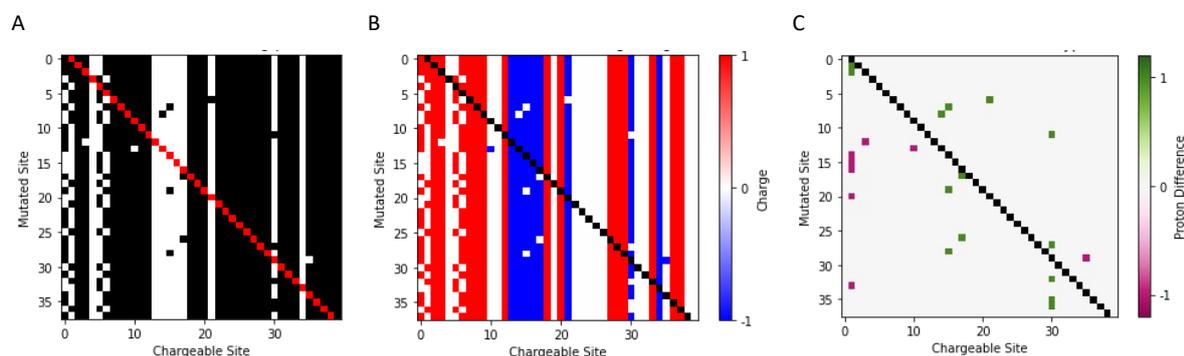
Figure S52 | Plots showing the alanine scanning functionality of ChargePlacer on PksACP4[8+], Coulomb-only mode.
A) Matrix plot of the proton pattern for each *in silico* mutant (rows), the mutated residues are indicated by red squares. B) Matrix plot of the charge pattern resulting from A, the mutated residues are indicated by black squares. C) Matrix plot comparing the proton patterns of the mutant (from A) against the nearest proton pattern for the 'wild-type' structure, as determined by replicates as shown in Figure S49. This shows that for the majority of *in silico* mutants, the effect on the proton pattern, and therefore charge distribution, is minimal. In all cases, at most a single proton is displaced, see Table S8 for details.

Table S7 | Change in $E_{Coulomb}$ (in kJ/mol) for *in silico* PksACP4[8+] mutants to alanine, Coulomb-only mode.
These mutants, shown in Figure S52, have proton patterns that differ from either of the two wild-type proton patterns. Mutants not present in this table had patterns that matched one of the wild-type patterns and showed no difference in E_{Coul}. Residues H−29 and E61 are prominent victims for the loss of negatively and positively chargeable residues, respectively.

Residue	+H	−H	$\Delta E_{Coulomb}$	Residue	+H	−H	$\Delta E_{Coulomb}$	Residue	+H	−H	$\Delta E_{Coulomb}$
H−28	−29		103.3	D20		−29	25.4	K59	20		446.9
H−27	−29		43.2	E24		−29	70.8	E61		88	50.4
R−17	31		463.8	E25		−29	118.2	R77	61		355.2
H−14	20		339.9	R27	25		447.9	D80		−29	63.2
R−2	13		495.7	R30	20		200.9	R89	61		328.1
K6	61		504.2	E31		−29	221.6	R97	61		429.5
D8		−27	535.1	R55	25		461.3				
E13		2	257.5	R58	61		298.9				

Gas-Phase Simulations of PksJ ACP4

Table S8 | Proton and charge placement for gas-phase simulations of PksACP4.
From the same structure, the 8+ and 7+ energy minimised proton patterns were determined accounting for both $E_{Coulomb}$ and E_{PA}. For the 8+ charge state, a set of variants were produced with Coulomb-only energy calculation (V1, V4 and V5) or Coulomb-PA energy calculation (V2 and V3). These variants were selected as the next best proton pattern for the determined score, or best in the case of V1.

		E_{tot} 8+		E_{tot} 7+		**Variant 1**		Variant 2		Variant 3		Variant 4		Variant 5	
# -ve		2		2		8		2		2		8		8	
# +ve		10		9		16		10		10		16		16	
Minimisation		Coulomb-PA		Coulomb-PA		Coulomb only		Coulomb-PA		Coulomb-PA		Coulomb only		Coulomb only	
$E_{Coulomb}$		657 kJ/mol		115 kJ/mol		-1035 kJ/mol		744 kJ/mol		707 kJ/mol		-1030 kJ/mol		-991 kJ/mol	
E_{PA}		37401 kJ/mol		36399 kJ/mol		34425 kJ/mol		37401 kJ/mol		37357 kJ/mol		34496 kJ/mol		34425 kJ/mol	
E_{tot}		-36744 kJ/mol		-36284 kJ/mol		-35460 kJ/mol		-36657 kJ/mol		-36650 kJ/mol		-35526 kJ/mol		-35416 kJ/mol	
Resn	**Resi**	z	H^+	z	H^+	z	H^+	z	H^+	z	H^+	z	H^+	z	H^+
NT	-32	0	FALSE	0	FALSE	**1**	**TRUE**	0	FALSE	0	FALSE	0	FALSE	1	TRUE
HIS	-29	0	FALSE	0	FALSE	0	FALSE	0	FALSE	0	FALSE	**1**	**TRUE**	1	TRUE
HIS	-28	0	FALSE	0	FALSE	1	TRUE	0	FALSE	1	TRUE	1	TRUE	1	TRUE
HIS	-27	1	TRUE	1	TRUE	1	TRUE	0	FALSE	0	FALSE	1	TRUE	1	TRUE
HIS	-26	0	FALSE	0	FALSE	0	FALSE	0	FALSE	0	FALSE	0	FALSE	0	FALSE
HIS	-25	0	FALSE	0	FALSE	0	FALSE	0	FALSE	0	FALSE	**1**	**TRUE**	0	FALSE
HIS	-24	0	FALSE	0	FALSE	**1**	**TRUE**	1	TRUE	1	TRUE	0	FALSE	0	FALSE
ARG	-17	1	TRUE	1	TRUE	1	TRUE	1	TRUE	1	TRUE	1	TRUE	1	TRUE
HIS	-14	1	TRUE	1	TRUE	1	TRUE	1	TRUE	1	TRUE	1	TRUE	1	TRUE
ARG	-2	1	TRUE	1	TRUE	1	TRUE	1	TRUE	1	TRUE	1	TRUE	1	TRUE
ASP	2	0	TRUE	0	TRUE	0	TRUE	0	TRUE	0	TRUE	0	TRUE	0	TRUE
GLU	3	0	TRUE	0	TRUE	0	TRUE	0	TRUE	0	TRUE	0	TRUE	0	TRUE
LYS	6	1	TRUE	1	TRUE	1	TRUE	1	TRUE	1	TRUE	1	TRUE	1	TRUE
ASP	8	-1	FALSE	-1	FALSE	-1	FALSE	-1	FALSE	-1	FALSE	-1	FALSE	-1	FALSE
GLU	13	-1	FALSE	-1	FALSE	-1	FALSE	-1	FALSE	-1	FALSE	-1	FALSE	-1	FALSE
ASP	20	0	TRUE	0	TRUE	-1	FALSE	0	TRUE	0	TRUE	-1	FALSE	-1	FALSE
GLU	24	0	TRUE	0	TRUE	-1	FALSE	0	TRUE	0	TRUE	-1	FALSE	-1	FALSE
GLU	25	0	TRUE	0	TRUE	-1	FALSE	0	TRUE	0	TRUE	-1	FALSE	-1	FALSE
ARG	27	1	TRUE	1	TRUE	1	TRUE	1	TRUE	1	TRUE	1	TRUE	1	TRUE
ASP	29	0	TRUE	0	TRUE	0	TRUE	0	TRUE	0	TRUE	0	TRUE	0	TRUE
ARG	30	0	FALSE	0	FALSE	1	TRUE	0	FALSE	0	FALSE	1	TRUE	1	TRUE
GLU	31	0	TRUE	0	TRUE	-1	FALSE	0	TRUE	0	TRUE	-1	FALSE	-1	FALSE
ASP	32	0	TRUE	0	TRUE	0	TRUE	0	TRUE	0	TRUE	0	TRUE	0	TRUE
GLU	34	0	TRUE	0	TRUE	0	TRUE	0	TRUE	0	TRUE	0	TRUE	0	TRUE
ASP	36	0	TRUE	0	TRUE	0	TRUE	0	TRUE	0	TRUE	0	TRUE	0	TRUE
ASP	41	0	TRUE	0	TRUE	0	TRUE	0	TRUE	0	TRUE	0	TRUE	0	TRUE
ASP	45	0	TRUE	0	TRUE	0	TRUE	0	TRUE	0	TRUE	0	TRUE	0	TRUE
ARG	55	0	FALSE	0	FALSE	1	TRUE	0	FALSE	0	FALSE	1	TRUE	1	TRUE
ARG	58	1	TRUE	0	FALSE	1	TRUE	1	TRUE	0	FALSE	1	TRUE	1	TRUE
LYS	59	0	FALSE	0	FALSE	1	TRUE	0	FALSE	0	FALSE	1	TRUE	1	TRUE
GLU	61	0	TRUE	0	TRUE	-1	FALSE	0	TRUE	0	TRUE	-1	FALSE	-1	FALSE
ASP	65	0	TRUE	0	TRUE	0	TRUE	0	TRUE	0	TRUE	0	TRUE	0	TRUE
GLU	71	0	TRUE	0	TRUE	0	TRUE	0	TRUE	0	TRUE	0	TRUE	0	TRUE
ARG	77	1	TRUE	1	TRUE	1	TRUE	1	TRUE	1	TRUE	1	TRUE	1	TRUE
ASP	80	0	TRUE	0	TRUE	-1	FALSE	0	TRUE	0	TRUE	-1	FALSE	-1	FALSE
GLU	88	0	TRUE	0	TRUE	0	TRUE	0	TRUE	0	TRUE	0	TRUE	0	TRUE
ARG	89	1	TRUE	1	TRUE	1	TRUE	1	TRUE	1	TRUE	1	TRUE	1	TRUE
ARG	97	1	TRUE	1	TRUE	1	TRUE	1	TRUE	1	TRUE	1	TRUE	1	TRUE
ASP	100	0	TRUE	0	TRUE	0	TRUE	0	TRUE	0	TRUE	0	TRUE	0	TRUE
CT	104	0	TRUE	0	TRUE	0	TRUE	0	TRUE	0	TRUE	0	TRUE	0	TRUE
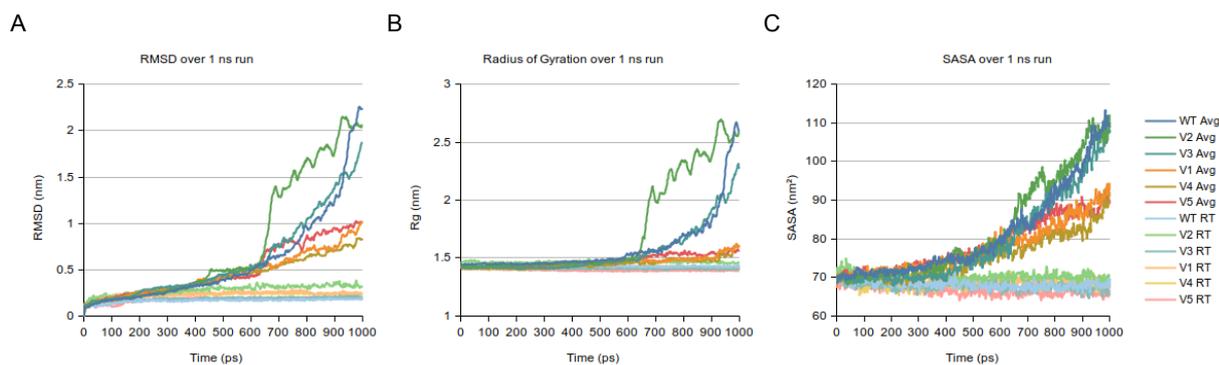
Figure S53 | Analysis of gas-phase simulations for proton pattern variants (as in Table S8) of the wild-type PksACP4. Standard analyses using built-in GROMACS programs for A) RMSD, B) Radius of Gyration and C) SASA. Data are shown for simulations at room temperature (298 K, **RT**) and for linear thermal gradients (298–798 K); for the latter, averages of 3 independent replicates are shown. This shows that Coulomb-only variants (V1, V4 and V5) are more stable during the thermal gradient simulations than the Coulomb-PA variants (WT, V2 and V3). At 298 K, there is little difference between any of the variants.
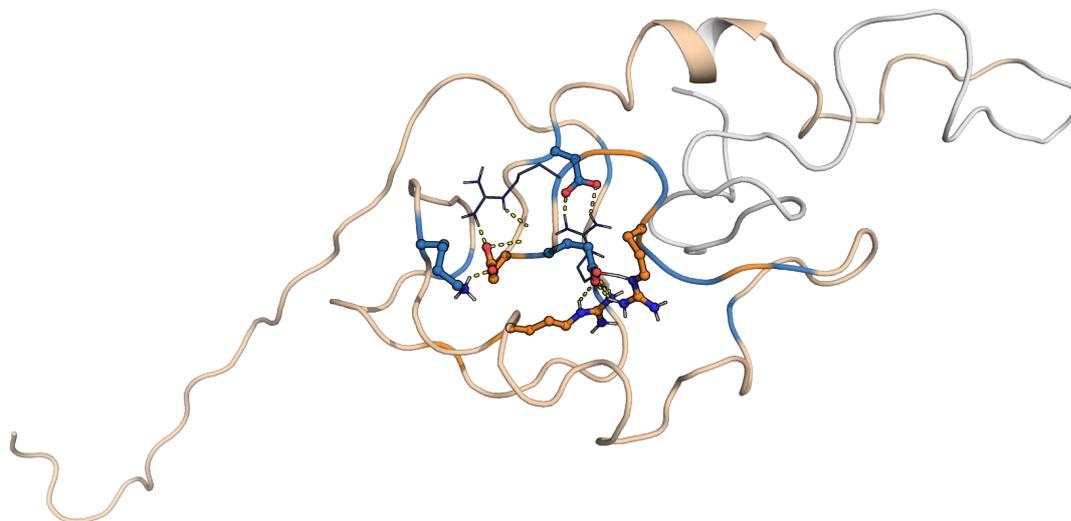
Figure S54 | Extended unfolded conformation of PksACP4^{8+} after thermal activation MD simulations for 5 ns. Cartoon rendering of structure after 1 ns thermal gradient (298–798 K) and a further 4 ns held at 798 K. Main chain is shown in wheat, the N-terminal tag is shown in grey. Residue sidechains for D20, E24, E25, R30, R55 and K59 are shown as sticks; R27 and R77 as lines. Residues are coloured according to ΔCIU50$_{IWAT}$: ≥1σ, orange; ≥2σ, blue. Polar contacts (as detected in PyMol) are shown as dashed lines.
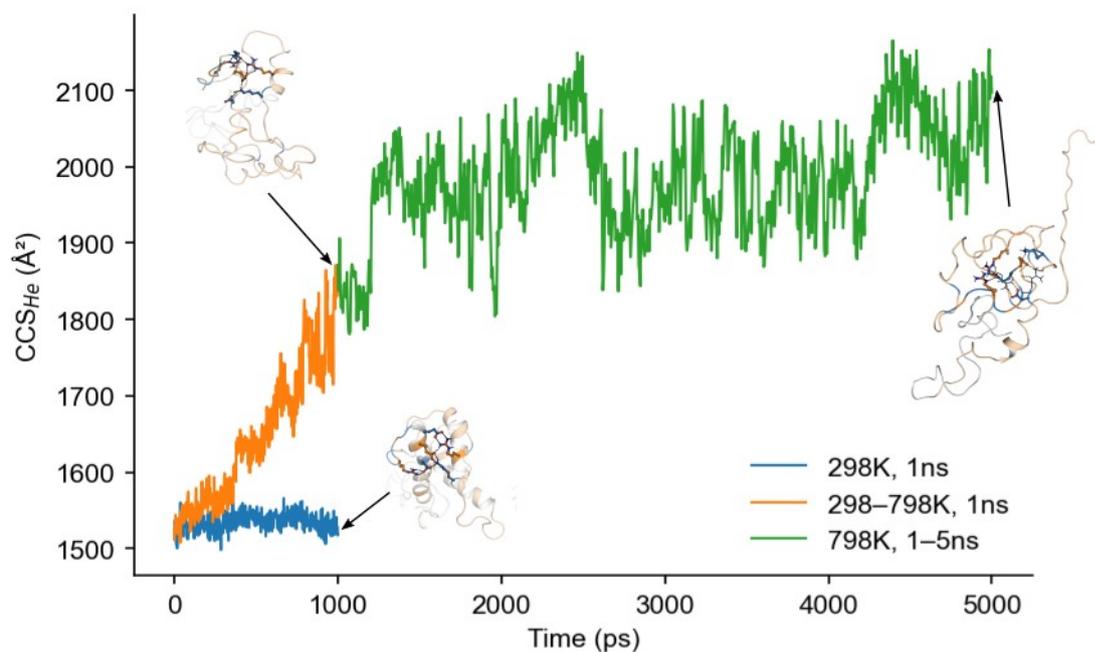
Figure S55 | Plot of CCS calculated from gas-phase MD simulations over time for PksACP4[8+].
CCS_{calc} values for structures during gas-phase MD simulations at 298 K for 1 ns (blue), thermal gradient over 298–798 K for 1 ns (orange) and additional 4 ns held at 798 K (green). Structures for the end points of each of these are shown, corresponding to structures shown in Figure 5A and Figure S54, respectively.
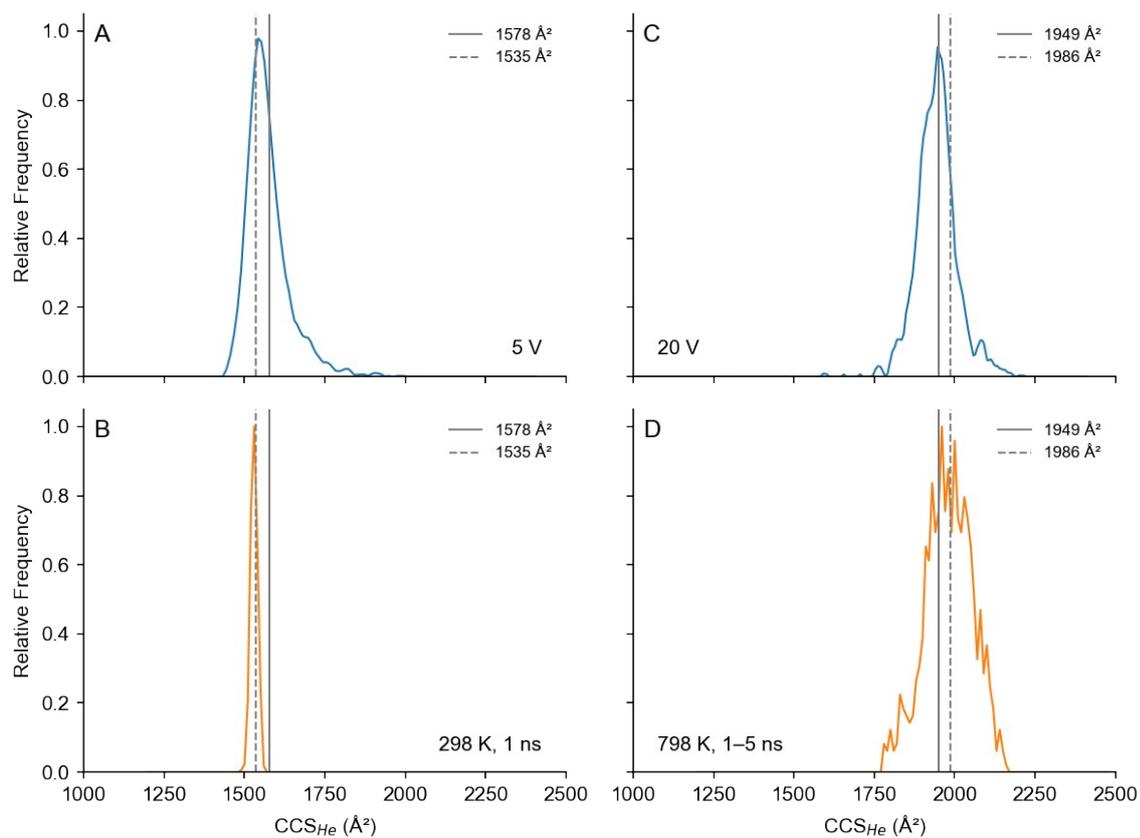
Figure S56 | PksACP4[8+] IMS data (A and C) and gas-phase MD data (B and D) transformed into CCS_{He} reference frame. A) $^{TW}CCS_{N2 \to He}$ for compact folded PksACP4[8+] at 5 V (40 eV). B) CCS_{calc} for all structures in 1 ns 298 K run. C) $^{TW}CCS_{N2 \to He}$ for extended unfolded PksACP4[8+] at 20 V (160 eV). D) CCS_{calc} for all structures in post-thermal gradient 1–5 ns 798 K run. Weighted-mean values for $^{TW}CCS_{N2 \to He}$ (solid line) and CCS_{calc} (dashed line) for the compact conformer (A and B) and the extended conformer (C and D) show very good agreement between IMS and MD data.

Bioinformatics
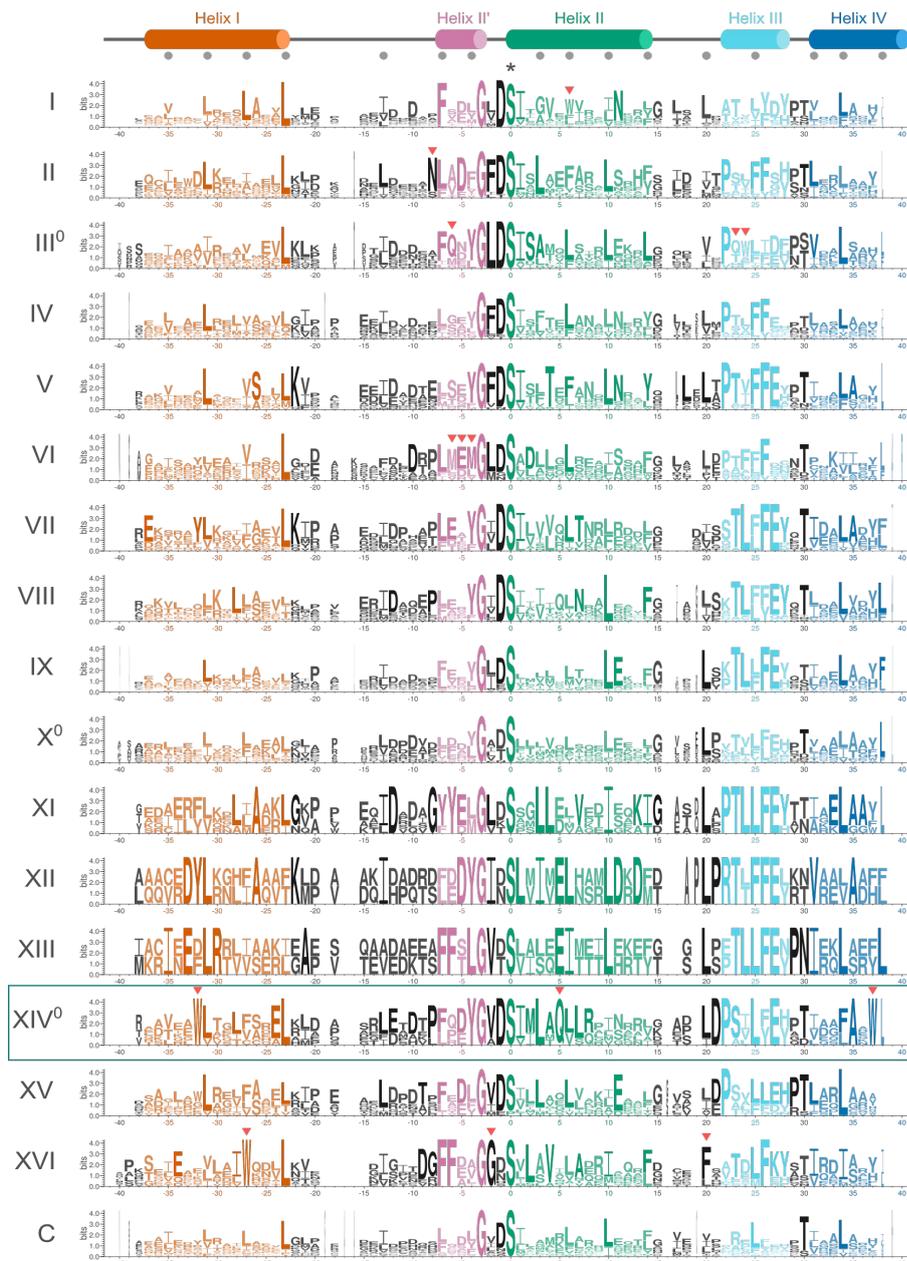
Figure S57 | Sequence logos for ACPs grouped by downstream KS clade.
Sequences were taken from an alignment of 313 ACPs, without RhiACP8 and LkcACP2 as these introduced two large sections of gaps between helices II and III. Logos were generated using webLogo v3.7.4,[21] and manually coloured to reflect helix positions and conservation. The conserved serine for Ppant attachment is highlighted with an *, and all positions are relative to this. Potentially important residues are highlighted with a triangle. Helices are derived from a solution NMR structure of MmpACP7a (PDB: 2L22, 1–76)[20] and the positions of residues in the hydrophobic core indicated with circles.
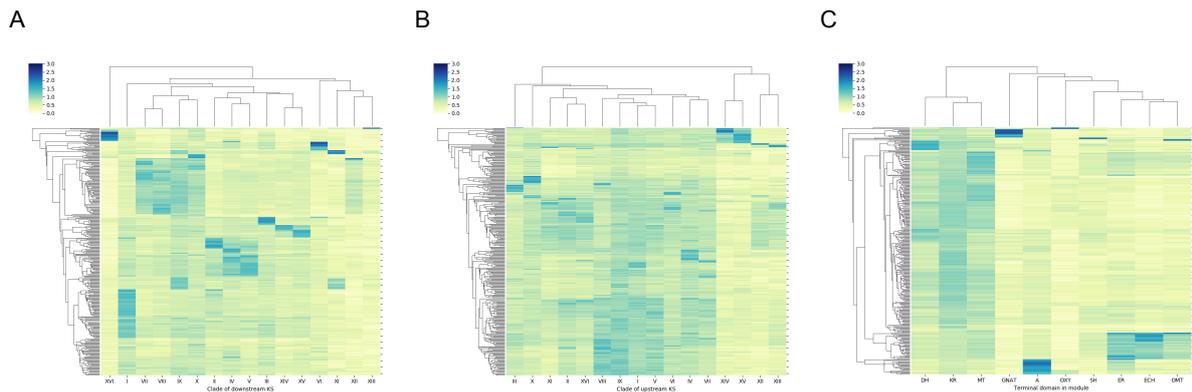
Figure S58 | Clustered heat-maps of HMM score ratios between grouped ACPs and all ACPs. Generated using the python package `seaborn` with the `clustermap` function, clustered in bow rows and columns. A value of >1 (green→blue) implies a greater match to grouped assignments.
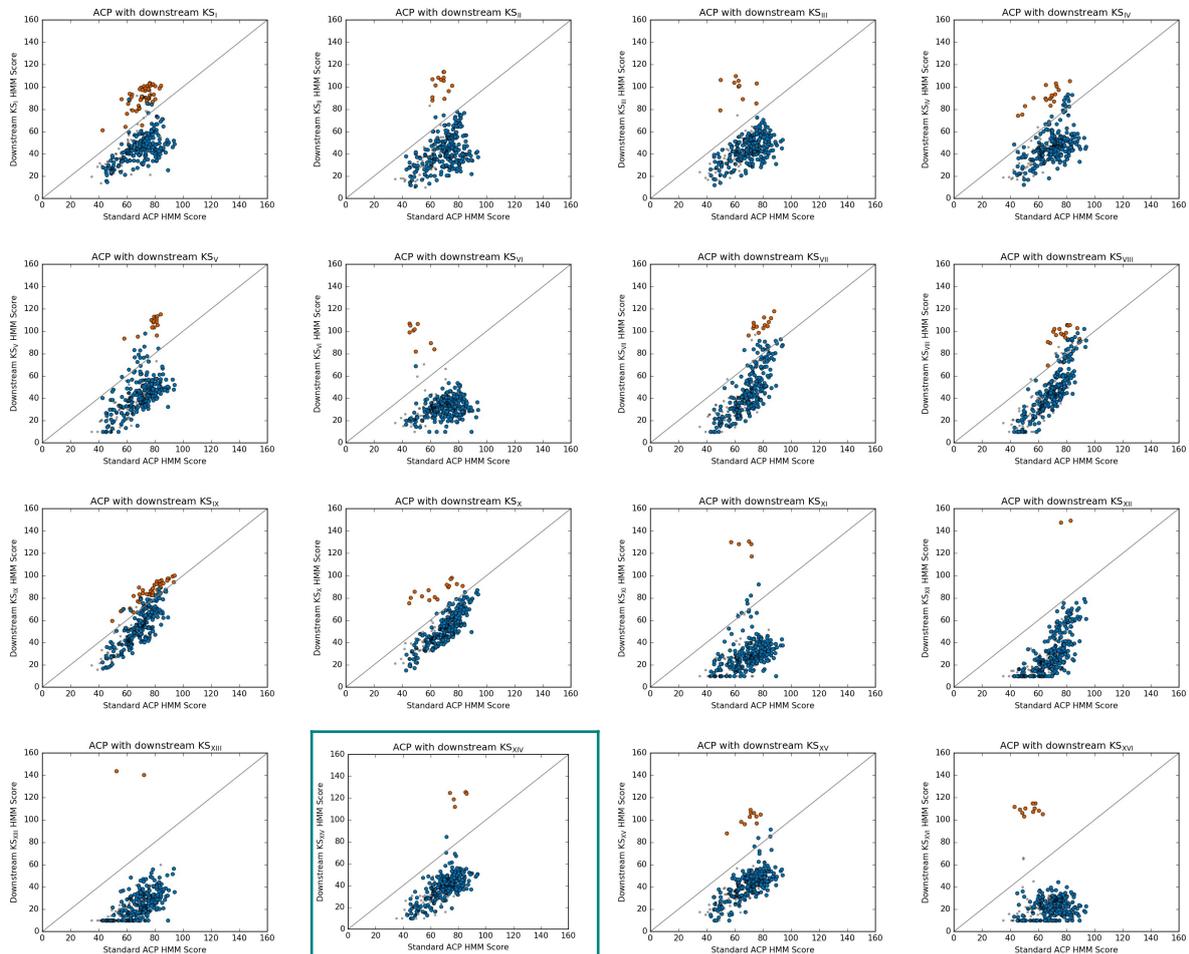
Figure S59 | Plots of HMM bit scores for ACPs grouped by **downstream** KS clade against all ACPs tested. ACPs from the group used to generate each profile are shown as orange dots. ACPs from other groups are shown as blue dots. ACPs with no assigned group are shown as grey dots (n=73).
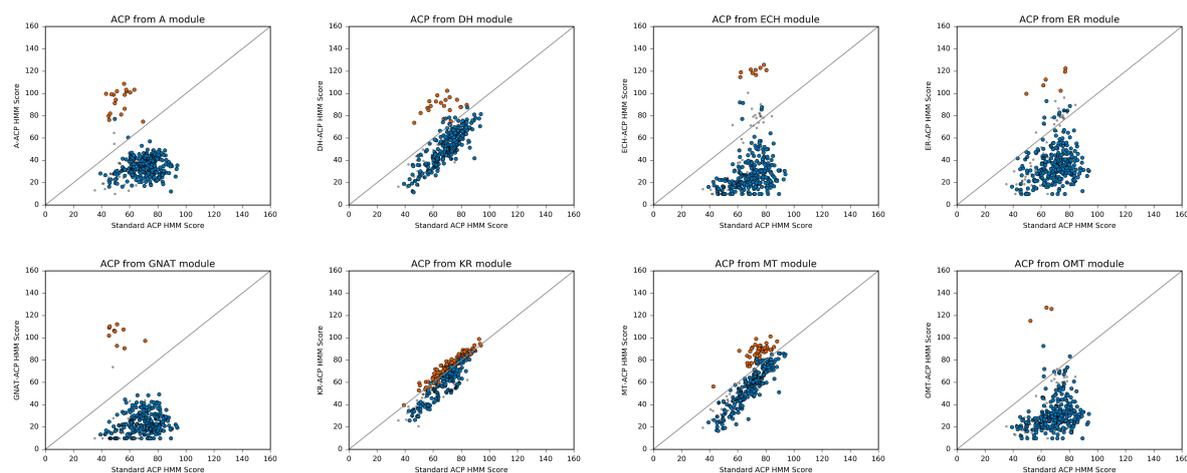
Figure S60 | Plot of HMM bit scores for ACPs grouped by **module-terminating-domain** against all ACPs tested. ACPs from the group used to generate each profile are shown as orange dots. ACPs from other groups are shown as blue dots. ACPs with no assigned group are shown as grey dots (n=64).

References

[1]   P. D. Straight, M. A. Fischbach, C. T. Walsh, D. Z. Rudner, R. Kolter, *Proc. Natl. Acad. Sci.* **2007**, *104*, 305–310.

[2]   J. D. Eschweiler, J. N. Rabuck-Gibbons, Y. Tian, B. T. Ruotolo, *Anal. Chem.* **2015**, *87*, 11516–11522.

[3]   D. A. Polasky, S. M. Dixit, S. M. Fantin, B. T. Ruotolo, *Anal. Chem.* **2019**, *91*, 3147–3155.

[4]   D. E. Clemmer, "Clemmer Collision Cross Section Database," can be found under https://clemlab.sitehost.iu.edu/Research/Cross%20Section%20Database/cs_database.php, **2020**.

[5]   M. F. Bush, "Bush Collision Cross Section Database," can be found under https://depts.washington.edu/bushlab/ccsdatabase/, **2020**.

[6]   M. F. Bush, Z. Hall, K. Giles, J. Hoyes, C. V. Robinson, B. T. Ruotolo, *Anal. Chem.* **2010**, *82*, 9557–9565.

[7]   R. Salbo, M. F. Bush, H. Naver, I. Campuzano, C. V. Robinson, I. Pettersson, T. J. D. Jørgensen, K. F. Haselmann, *Rapid Commun. Mass Spectrom.* **2012**, *26*, 1181–1193.

[8]   B. T. Ruotolo, J. L. P. Benesch, A. M. Sandercock, S.-J. Hyung, C. V. Robinson, *Nat. Protoc.* **2008**, *3*, 1139–1152.

[9]   D. V. D. Spoel, E. Lindahl, B. Hess, G. Groenhof, A. E. Mark, H. J. C. Berendsen, *J. Comput. Chem.* **2005**, *26*, 1701–1718.

[10]   G. A. Kaminski, R. A. Friesner, J. Tirado-Rives, W. L. Jorgensen, *J. Phys. Chem. B* **2001**, *105*, 6474–6487.

[11]   A. Moser, K. Range, D. M. York, *J. Phys. Chem. B* **2010**, *114*, 13911–13921.

[12]   L. Konermann, H. Metwally, R. G. McAllister, V. Popa, *Methods* **2018**, *144*, 104–112.

[13]   J. L. Benesch, B. T. Ruotolo, *Curr. Opin. Struct. Biol.* **2011**, *21*, 641–649.

[14]   S. C. Epstein, L. K. Charkoudian, M. H. Medema, *Stand. Genomic Sci.* **2018**, *13*, 16.

[15]   F. Sievers, A. Wilm, D. Dineen, T. J. Gibson, K. Karplus, W. Li, R. Lopez, H. McWilliam, M. Remmert, J. Söding, J. D. Thompson, D. G. Higgins, *Mol. Syst. Biol.* **2011**, *7*, 539.

[16]   F. Sievers, D. G. Higgins, in *Mult. Seq. Alignment Methods* (Ed.: D.J. Russell), Humana Press, Totowa, NJ, **2014**, pp. 105–116.

[17]   S. R. Eddy, *PLOS Comput. Biol.* **2008**, *4*, e1000069.

[18]   S. R. Eddy, *PLOS Comput. Biol.* **2011**, *7*, e1002195.

[19]   T. Nguyen, K. Ishida, H. Jenke-Kodama, E. Dittmann, C. Gurgui, T. Hochmuth, S. Taudien, M. Platzer, C. Hertweck, J. Piel, *Nat. Biotechnol.* **2008**, *26*, 225–233.

[20]   A. S. Haines, X. Dong, Z. Song, R. Farmer, C. Williams, J. Hothersall, E. Płoskoń, P. Wattana-amorn, E. R. Stephens, E. Yamada, R. Gurney, Y. Takebayashi, J. Masschelein, R. J. Cox, R. Lavigne, C. L. Willis, T. J. Simpson, J. Crosby, P. J. Winn, C. M. Thomas, M. P. Crump, *Nat. Chem. Biol.* **2013**, *9*, 685–692.

[21]   G. E. Crooks, G. Hon, J.-M. Chandonia, S. E. Brenner, *Genome Res.* **2004**, *14*, 1188–1190.