current issues in molecular biology

MDPI

Supplementary Information

From Scar-Free Skin Repair in Leopard Geckos to Lig-and-Dependent Reorganization at the Human KEAP1–NRF2 Interface

Yian Li [1], Jianlin Li [1], Dong-Qing Wei [1], Wei Miao [2,*] and Jiayi Li [1,*]

[1] Joint International Research Laboratory of Metabolic & Developmental Sciences, School of Life Sciences and Biotechnology, Shanghai Jiao Tong University, Shanghai 200040, China
[2] McCormick School of Engineering, Northwestern University, Evanston, IL 60208, USA
* Correspondence: drmiaowei@163.com (W.M.); jemmylee@sjtu.edu.cn (J.L.)

Supplementary Methods; Tables S1–S5; Figures S1–S12.

Supplementary Methods

This supplement provides simulation inputs, analysis definitions, screening procedures and sequence-comparison methods for the human KEAP1 study. Tables S1–S5 and Figures S1–S12 contain the associated chemical identities, quantitative summaries and supporting analyses. Reference numbers correspond to the main manuscript.

SM1. Structural inputs and simulation preparation

The principal dataset comprises WT and Q530A human KEAP1 (UniProt Q14145, residues 325–609) in apo, C1-, C2- and C3-bound states. Each condition has three 200 ns trajectories initialized with different heating velocity seeds, sharing the prepared coordinates and topology within that condition. All three selected candidate poses used receptor B1_10, taken at 47.31 ns from macrocycle-63 reference trajectory 3 (L_63_r3). B1_10 was selected as the actual frame nearest a receptor-feature cluster center in the ensemble construction described in SM5. WT apo was generated by ligand removal from this conformation; Q530A was introduced into the corresponding apo and holo starts.

Structural inputs were PDB 6TYM–08A, 2FLU–NRF2 peptide, 7Q8R–63, 7Q96–64 and 4IQK–IQK [5,7,8,13,15]. The labels 08A and IQK identify PDB chemical components; 63 and 64 are publication compound numbers. Known-ligand poses were transferred to a common 6TYM-derived receptor. The peptide model contains NRF2 residues 69–84 (AFFAQLQLDEETGEFL), including the rebuilt E78 side chain, with free charged termini. C1, C2 and C3 correspond to source identifiers NSC2518124, NSC2530227 and NSC2532809, respectively; their exact simulated chemical states are specified in Table S1.

Protein and peptide parameters used ff14SB [16]. Ligands used GAFF2 atom types, AM1-BCC charges and supplementary parameters assigned by parmchk2 [17,18]. Histidines were HIE, cysteines were reduced CYS, and truncated protein termini were charged. WT protein net charge was −6 e; the peptide was −4 e. Reference ligands 08A, 63 and 64 were monoanions, while IQK was neutral. C1 and C3 were monoanions. C2 was modeled as a dianion with deprotonated carboxylate and sulfonamide nitrogen.

Systems were solvated in rectangular TIP3P boxes with a 12 Å buffer, neutralized and supplemented with approximately 0.15 M NaCl pairs. LEaP loaded leaprc.water.tip3p, including frcmod.tip3p and the TIP3P-specific ion parameter files. The Na+/Cl− parameters were supplied by frcmod.ionsjc_tip3p (Rmin/2 = 1.369/2.513 Å; ε = 0.0874393/0.0355910 kcal mol−1, respectively). Preparation comprised restrained minimization and unrestrained minimization (each up to 5,000 cycles, switching from steepest descent after 2,500), 100 ps NVT heating from 10 to 298.15 K, 500 ps NPT density relaxation and 1 ns unrestrained NPT equilibration. Positional-restraint coefficients were 5, 0, 2,

1 and 0 kcal mol−1 Å−2, respectively. Protein backbone atoms were restrained; ligand heavy atoms were additionally restrained in small-molecule systems. The reference peptide itself was unrestrained.

Production used Amber 26 GPU PMEMD, 2 fs integration, SHAKE on bonds involving hydrogen (ntc = 2; ntf = 2), particle-mesh Ewald electrostatics, a 9 Å direct-space cutoff and Langevin temperature control with collision frequency 2 ps−1. Isotropic Monte Carlo pressure control maintained 1 bar. Coordinates were saved every 10 ps. Production time excludes 1.6 ns of heating, density relaxation and equilibration. The 100–200 ns runs continued the original trajectories.

SM2. Structural observables and spatial coverage

After imaging with CPPTRAJ [19], production trajectories were sampled every 100 ps. All 285 protein Cα atoms were fitted to the common 6TYM-derived receptor frame. Protein Cα RMSD was calculated relative to the prepared WT apo structure in that frame. Figures S1 and S2 display unsmoothed 0.01–199.91 ns observations with separate colors for individual trajectories. The primary 20–200 ns window contains 1,800 sampled frames per trajectory; each 80 ns sensitivity window contains 800.

Protein radius of gyration was calculated from all protein atoms and topology masses as $Rg = [\Sigma_i m_i | r_i - rCOM|^2/\Sigma_i m_i]^{1/2}$. Ligand, water and ions were excluded. Rg is invariant to rigid-body alignment. Cα RMSF describes fluctuations around each trajectory's own mean coordinates in the common fit.

The interface mask was fixed from the union of receptor heavy-atom contacts within 6 Å of prepared 08A and aligned ETGE79–82. It comprises residues 333, 334, 338, 363–365, 380, 382, 414–416, 461–463, 477, 478, 483, 508, 509, 524–527, 530, 555, 556, 572, 574, 577 and 602–604. Common-pocket RMSD uses 229 heavy-atom identities shared by WT and Q530A. Arginine guanidinium centers are the unweighted centroids of CZ, NH1 and NH2; R415–R483 distance uses these centers. For each pocket observable, the relative geometric contrast was calculated as (Q530A holo − WT holo) − (Q530A apo − WT apo) from the equally weighted group means.

Ligand heavy-atom RMSD was measured in the receptor frame relative to the prepared ligand pose; self-fitted RMSD measured internal deformation. Ring-core atoms belong to a graph cycle, and non-ring atoms outside the carboxylate form the peripheral set. Carboxylate displacement is the C/O/O centroid shift from the prepared pose. For the NRF2 peptide, ETGE residues 79–82 define the core.

ETGE reference-site coverage is the fraction of the 29 reference heavy-atom positions within 3 Å of at least one ligand heavy atom. The distal descriptor uses all 13 G81/E82 sites together (4 G81 and 9 E82 atoms), with equal weight per atom. Whole-ETGE sensitivity calculations used 2 and 4 Å cutoffs; global versus interface-Cα fitting; and alternative peptide conformations A–C, selected as the actual frames nearest each peptide trajectory's ETGE coordinate mean. Atom-set checks excluded ligand carboxylate points, reference E79 points or both. These Figure S5 comparisons assess the whole-ETGE descriptor and are distinct from the targeted R483-conditioned distal-coverage comparisons in Figure S8.

Ligand solvent-accessible surface area used the Shrake–Rupley implementation in MDTraj [20], a 1.4 Å probe and 960 surface points per atom at 1 ns spacing. All solute atoms blocked accessibility, but only ligand areas were summed. The isolated area used the same ligand conformation without protein. Buried fraction equals one minus the bound-to-isolated area ratio.

SM3. Hydrogen bonds, contact states and local hydration

Chemically typed N/O hydrogen bonds required donor–acceptor distance ≤3.5 Å and donor–hydrogen–acceptor angle ≥135°. Per-residue occupancy was the fraction of frames with any qualifying event; multiple atom pairs or donor

hydrogens in one frame were counted once. A water bridge required one water molecule to satisfy both protein–water and water–ligand criteria in the same sampled frame. Contact events were evaluated at 100 ps resolution.

R415/R483 contact categories were the same-frame union states: R483 only, R415 only, both and neither. C2 ring-nitrogen and deprotonated sulfonamide-nitrogen events were counted separately and jointly; "Both N groups" required simultaneous contacts to R415. Individual functional-group occupancies included overlapping events. Chemical identities and simulated microstates are specified in Table S1.

Conditional coverage used frames with any typed R483 direct hydrogen bond, with R483–carboxylate hydrogen bonding as a stricter check. WT C3 was separately conditioned on Q530 direct hydrogen bonding. Temporal contact runs were contiguous sampled frames with the same exclusive-contact label; alternating R415-only and R483-only runs were counted with minimum durations of 0.1, 1 and 5 ns, allowing intervening Both/Neither states. These summaries counted changes between sampled contact labels.

R415/R483 side-chain axes ran from Cα to Cζ and were compared with the prepared WT apo reference after the common receptor fit. χ1–χ4 angles were retained in the analysis records. The ligand ring axis ran from the C/O/O carboxylate centroid to the centroid of all ring heavy atoms; its angle to the vector toward the fixed G81/E82 reference centroid described distal-facing direction.

Solvent analysis sampled 0.01–199.01 ns every 1 ns from the original periodic trajectories. Counts used water oxygen atoms within 4 Å of residue-530 Cβ, with 5 Å as a radius check, and within 3.5 Å of the Cζ/Nη1/Nη2 guanidinium centroid of R415 or R483. Orthorhombic minimum-image distances and molecule-consistent centroids were used. The principal window contributed 180 solvent observations per trajectory. Counts include the change in accessible volume associated with the residue-530 side chain.

C2 structures in Figures 4a,b and 5b use WT R483-only and Q530A R415-only frames, respectively. C1/C3 structures in Figure 5a,c and Figure S12 use R483-contacting frames. Figure 5d–f statistics use R483-present frames for all three ligands, so the C2 structural example and conditional distribution have different selection criteria. For each illustration, qualifying ligand coordinates were averaged within each trajectory; trajectory means then received equal weights, and the saved frame nearest the resulting group mean supplied the representative. Apo structures used the same procedure for common pocket heavy atoms over 20–200 ns, including only N/Cα/C/O/Cβ at residue 530. Topology-defined bonds were retained. PyMOL used a common orthoscopic camera, a pale receptor and a gray ETGE reference; WT/Q530A carbon atoms are blue/orange in matched views. Dashed contacts reproduce chemically typed hydrogen bonds in the selected frame. Overlaid models originate from separate simulations. Candidate chemical structures were drawn and exported in ChemDraw 26 from the recorded microstates; RDKit checked canonical isomeric identity and formal charge.

SM4. Endpoint energies and statistical summaries

MMPBSA.py with the AmberTools 26 Sander backend calculated matched PB and GB endpoint estimates [26]. Each holo trajectory contributed 180 snapshots at 1 ns spacing over 20.01–199.01 ns. Receptor and ligand coordinates were extracted from the corresponding complex snapshots. Calculations used mbondi2 radii, dielectric constants 1 and 80, and 0.15 M salt. PB settings were radiopt = 0, inp = 2, scale = 2.0, cavity_surften = 0.0378, cavity_offset = −0.5692 and fillratio = 4. GB settings were igb = 5, alpb = 0, surften = 0.0072 and surfoff = 0. Native residue decomposition used idecomp = 2 and gbsa = 2, whereas the GB endpoint total used LCPO for its nonpolar term. Entropy was omitted.

The principal window was 20–200 ns. Trajectories were the statistical units (n = 3 per condition) and received equal weights; means and sample SD were calculated from trajectory summaries. Conditional summaries first selected qualifying frames within each trajectory and then averaged the resulting trajectory means. Density and box summaries assigned equal total weight to each trajectory, including when conditional frame counts differed; weighted empirical

quartiles were obtained by interpolation of cumulative observation weights at their midpoints. Whiskers extend to observations within 1.5 interquartile ranges. Diamonds and error bars show means and between-trajectory SD. WT and Q530A are unpaired. Figure 3c summarizes within-trajectory 2 ns bins, whereas Figures S1 and S2 show unsmoothed individual observations.

Leave-one-trajectory-out diagnostics omitted one trajectory from its own group while retaining all three in the other group, yielding six contrasts. Equal-duration 20–100 and 120–200 ns windows assessed time dependence; 100–120 ns remained in the primary 20–200 ns window. Full-cohort conditional contrasts required qualifying frames from every trajectory. C3 R483-present observations occurred in every trajectory in each window. Under the stricter R483–carboxylate condition, full-window distal coverage was 0.554 ± 0.188 in WT and 0.095 ± 0.021 in Q530A; the late-window full-cohort contrast was undefined because one Q530A trajectory lacked qualifying frames. Q530-conditioned WT C3 results use the full window, for which all three trajectories contributed. Missing conditional values remained undefined. Statistical summaries were descriptive, without frame-level significance tests. Unconditional C2 distal-coverage and C3 R483-contact contrasts were more sensitive to trajectory omission than the principal C2 exclusive-contact and C3 R483-conditioned coverage contrasts (Figure S8). Tables S2–S4 provide the principal structural, energetic and contact-state summaries.

SM5. Separate reference simulations and docking evaluation

The reference cohort comprised seven WT conditions with three trajectories per condition: 6TYM-derived apo (A_S), 2FLU-derived apo (A_P), NRF2 peptide (P_ETGE), and ligands 08A, 63, 64 and IQK. Reference comparisons used the matched 20–100 ns segment from the original 100 ns production runs (21 trajectories; 1.68 μs analyzed in aggregate), separately from the 24 core trajectories. The first six conditions (18 trajectories) supplied the receptor ensemble; IQK was an additional ligand reference. Published SPR and ITC affinities of macrocycles 63 and 64 are shown separately with their reported SD [13].

The receptor/pharmacophore ensemble comprised 12 receptor frames and eight occupancy-weighted templates containing 50 features, fixed before library and external evaluation. The NCI NExT Diversity 3500 Pilot Set contained 3,560 records; after standardization and eligibility filtering, 2,042 parents were eligible and 2,036 were prepared, yielding 2,415 microstates and 2,433 conformers. Epik [21], ETKDGv3 [22] and AutoDock 4.2.6 [23] were used for state preparation, conformer generation and docking. Morgan radius-2, 2,048-bit fingerprints described similarity [24], alongside separate connectivity, stereochemical and microstate identity records.

Static docking used two searches per microstate; expanded static and ensemble searches each used 24, with ensemble searches distributed across 12 receptors. The ensemble pool was reranked with AD4 + 2(1 − C), where C is occupancy-weighted pharmacophore coverage. Development evaluation used 08A, 63, 64 and IQK. The eight external ligands were publication compounds 1, 3, 7, 23, 24, 29, 33 and 39 [10], represented by PDB entries 9HWQ, 9HWR, 9HWS, 9HWX, 9HWT, 9HWU, 9HWW and 9HWV, respectively. The external ligands formed one related lineage. Each compound used three search seeds. Recovery required whole-ligand symmetry-corrected RMSD ≤2 Å against the transferred experimental pose, without a separate ligand fit. ETGE reference-site coverage was added retrospectively to the complete 576-pose external ensemble pool without changing ranking or selection.

Library standardization retained the largest fragment and required elements H, C, N, O, F, P, S, Cl, Br or I, molecular weight 150–650, at most 12 rotatable bonds, absolute formal charge ≤2 and fully specified required tetrahedral and E/Z stereochemistry. Acyl halides, isocyanates, isothiocyanates, epoxides, acyclic Michael acceptors and quinones were excluded; PAINS patterns were flagged rather than excluded. LigPrep 2023-4 with Epik generated states at pH 7.4 ± 1.0. Rings larger than eight atoms received three graph-derived ETKDGv3 conformers with up to 2,000 embedding iterations; MMFF relaxation was attempted when parameters were available. Other states retained one conformer.

Meeko 0.8.0 assigned docking inputs with Gasteiger charges and explicit polar hydrogens, distinct from the MD charge model. Exact stereochemical graphs and charge sums were checked.

The docking grid used the transferred development-ligand bounds plus 6 Å per side: center (−6.2845, −23.2877, −19.9362) Å, point counts (66, 56, 60) and spacing 0.375 Å. Each Lamarckian genetic-algorithm search used population 150, up to 2.5 million evaluations and 27,000 generations, elitism 1, mutation rate 0.02, crossover rate 0.8 and local-search frequency 0.06. Solis–Wets searches allowed 300 iterations, success/failure limits of 4, initial rho 1 and lower rho 0.01. Translation, orientation and torsions were randomized. Macrocycle ring conformations were fixed within a docking search, and the per-state allowance was divided across prepared conformers.

Receptor selection used standardized R380–R483 and R415–R483 distances and sine/cosine components of R415, R483 and Y572 χ1 angles. Each of the 18 development trajectories contributed 800 frames at 100 ps spacing over 20–100 ns. A common scaler was fitted to the pooled eight-feature vectors; each of the six conditions was then partitioned separately into two k-means clusters (20 initializations; random seed 2611). The actual member nearest each cluster center supplied one of the 12 receptor frames. Pharmacophore templates required at least 80 observations and feature mean conditional occupancy ≥0.5. Occupancies, feature positions and directions assigned equal weights to contributing trajectories. Spatial tolerance was the larger of 1 Å and the weighted 90th-percentile positional deviation; direction tolerance was the larger of 30° and the corresponding angular deviation. Incoherent directions used a 180° cone. Pose matches also had to satisfy direct hydrogen-bond geometry. The maximum coverage among alternative templates defined C. An E82–R415 peptide template had support from one trajectory.

Each method under primary and Epik-adjusted ranking contributed its top 1% of parents to a 46-parent union. Additional search seeds evaluated the same union with all associated states and conformers. Final parent ranking used the median best score across seeds. The representative pose came from the seed nearest that median, with deterministic ties. C1–C3 were selected from distinct generic molecular cores under ensemble reranking before MD. External structures and published assay annotations were used for curation. The simulated C2 dianion was the state selected by the primary ensemble-reranking criterion (AD4 plus the pharmacophore penalty), which determined the pre-MD candidate pose. The separate Epik-adjusted sensitivity ranking favored a C2 monoanion.

Retrospective recovery results are shown in Figure S10. For static docking, expanded static search, ensemble docking and ensemble reranking, top-pose recovery was 3/12, 2/12, 0/12 and 0/12 development compound–seed cases, respectively; corresponding external counts were 9/24, 8/24, 10/24 and 11/24. Broad ETGE reference-site coverage and pharmacophore agreement also occurred among poses outside the 2 Å recovery criterion. The external evaluation used one related ligand lineage and three correlated search seeds per compound; external structures and assay annotations were available during curation.

SM6. Sequence sources and conservation

NCBI RefSeq proteins XP_054859680.1 (KEAP1) and XP_054826676.1 (NRF2) were retrieved with their full GenBank annotations [27]. Both were predicted full-length records from the leopard-gecko assembly GCF_028583425.1; neither contained ambiguous residues or internal stop characters. NCBI Datasets Ortholog assignments linked gecko KEAP1 Gene ID 129346357 to human Gene ID 9817, and gecko NFE2L2 Gene ID 129323889 to human Gene ID 4780. Reviewed human UniProtKB sequences were Q14145 and Q16236 [28]. Retrieval date was 4 October 2026.

Biopython global pairwise alignment used BLOSUM62, gap-opening score −10 and gap-extension score −0.5 [29]. Alternative penalties of −12/−0.5 and −10/−1 retained the mappings of all positions in human KEAP1 325–609 and NRF2 69–84. Identity was calculated over these fixed human-position sets and the predefined 32-residue interface mask. Site correspondence used the aligned human and gecko residue numbers. Figure S11 displays all corresponding positions

in the studied regions; Figure 1d highlights local neighborhoods. Equal-height residue letters and identity colors display the pairwise alignment.

Table S1. Candidate chemical identities and simulated states.

Candidate	Formula / charge	Simulated-state SMILES
C1	C22H16N3O2-	Cc1ccc(-c2nc(Nc3ccc(C(=O)[O-])cc3)c3ccccc3n2)cc1
C2	C21H14N4O4S-2	O=C([O-])c1cccc(Nc2nc3ccccc3nc2[N-]S(=O)(=O)c2ccccc2)c1
C3	C23H21N2O4-	O=C(Cn1nc(C(=O)[O-])c2ccccc2c1=O)c1ccc(C2CCCCC2)cc1

SMILES specify connectivity and charge for the simulated species. C1 and C3 are monoanions; C2 is a dianion. Source identifiers are C1, NSC2518124; C2, NSC2530227; and C3, NSC2532809.

Table S2. Principal structural summaries over 20–200 ns.

Condition	Ligand RMSD / Å	Self-fit RMSD / Å	Carboxylate shift / Å	ETGE coverage	G81/E82 coverage	Buried fraction
WT–C1	3.946 ± 1.447	1.018 ± 0.231	2.220 ± 0.303	0.718 ± 0.067	0.500 ± 0.075	0.769 ± 0.049
Q530A–C1	4.038 ± 0.877	1.121 ± 0.220	2.295 ± 0.256	0.703 ± 0.053	0.462 ± 0.057	0.781 ± 0.051
WT–C2	4.503 ± 1.351	1.765 ± 0.342	1.585 ± 0.647	0.853 ± 0.083	0.693 ± 0.171	0.735 ± 0.005
Q530A–C2	5.689 ± 1.255	1.958 ± 0.229	8.057 ± 2.794	0.717 ± 0.149	0.657 ± 0.166	0.663 ± 0.043
WT–C3	4.771 ± 1.635	2.214 ± 0.586	2.150 ± 0.504	0.723 ± 0.113	0.564 ± 0.179	0.667 ± 0.076
Q530A–C3	7.790 ± 0.973	2.993 ± 0.365	2.236 ± 0.564	0.487 ± 0.115	0.086 ± 0.056	0.652 ± 0.068

Values are equal-weight means of trajectory means ± between-trajectory SD. ETGE-site coverage uses a 3 Å cutoff; G81/E82 coverage denotes the distal reference region.

Table S3. Endpoint energy estimates over 20–200 ns (kcal/mol).

Ligand	WT PB	Q530A PB	WT GB	Q530A GB
C1	-7.721 ± 1.474	-7.217 ± 2.492	-36.686 ± 6.675	-37.483 ± 6.994
C2	-1.465 ± 4.918	-0.807 ± 0.893	-23.067 ± 1.577	-24.345 ± 1.363
C3	-5.843 ± 6.928	-7.246 ± 3.746	-30.068 ± 7.547	-33.544 ± 6.799

Values are group means ± between-trajectory SD. PB and GB estimates use matched snapshots and omit entropy.

Table S4. Contact-state and conditional-coverage summaries over 20–200 ns.

Observable	WT	Q530A
C2 R483 only fraction	0.861 ± 0.067	0.179 ± 0.202
C2 R415 only fraction	0.036 ± 0.054	0.627 ± 0.183
C2 R415: ring nitrogen	0.042 ± 0.034	0.598 ± 0.188
C2 R415: sulfonamide nitrogen	0.028 ± 0.046	0.618 ± 0.183
C2 R415: both nitrogen groups	0.022 ± 0.038	0.579 ± 0.180
C3 distal coverage: R483 contact	0.554 ± 0.188	0.084 ± 0.026
C3 distal coverage: R483–carboxylate	0.554 ± 0.188	0.095 ± 0.021

Values are equal-weight trajectory summaries ± between-trajectory SD. Functional-group contacts can overlap; the joint nitrogen row counts their intersection. Conditional coverage is calculated only in qualifying frames. The stricter late-window full-cohort C3 comparison was undefined because a trajectory had no qualifying frames.

Table S5. Human and leopard-gecko sequence correspondence.

Protein / region	Human	Leopard gecko	Comparison
KEAP1 sequence	Q14145; 624 aa	XP_054859680.1; 599 aa	Predicted RefSeq ortholog
Modeled KEAP1 segment	325–609	293–577	243/285 identical (85.3%)
KEAP1 interface	32 predefined sites	Corresponding positions	31/32 identical; F478/Y446 differs
Principal KEAP1 sites	R415; R483; Q530	R383; R451; Q498	All conserved
NRF2 sequence	Q16236; 605 aa	XP_054826676.1; 591 aa	Predicted RefSeq ortholog
NRF2 reference region	69–84	59–74	13/16 identical
NRF2 motif	ETGE79–82	ETGE69–72	4/4 identical

Sources: NCBI RefSeq and reviewed UniProtKB records retrieved on 4 October 2026 [27,28]. NCBI Ortholog assignments support the correspondence; gecko records are predicted proteins.

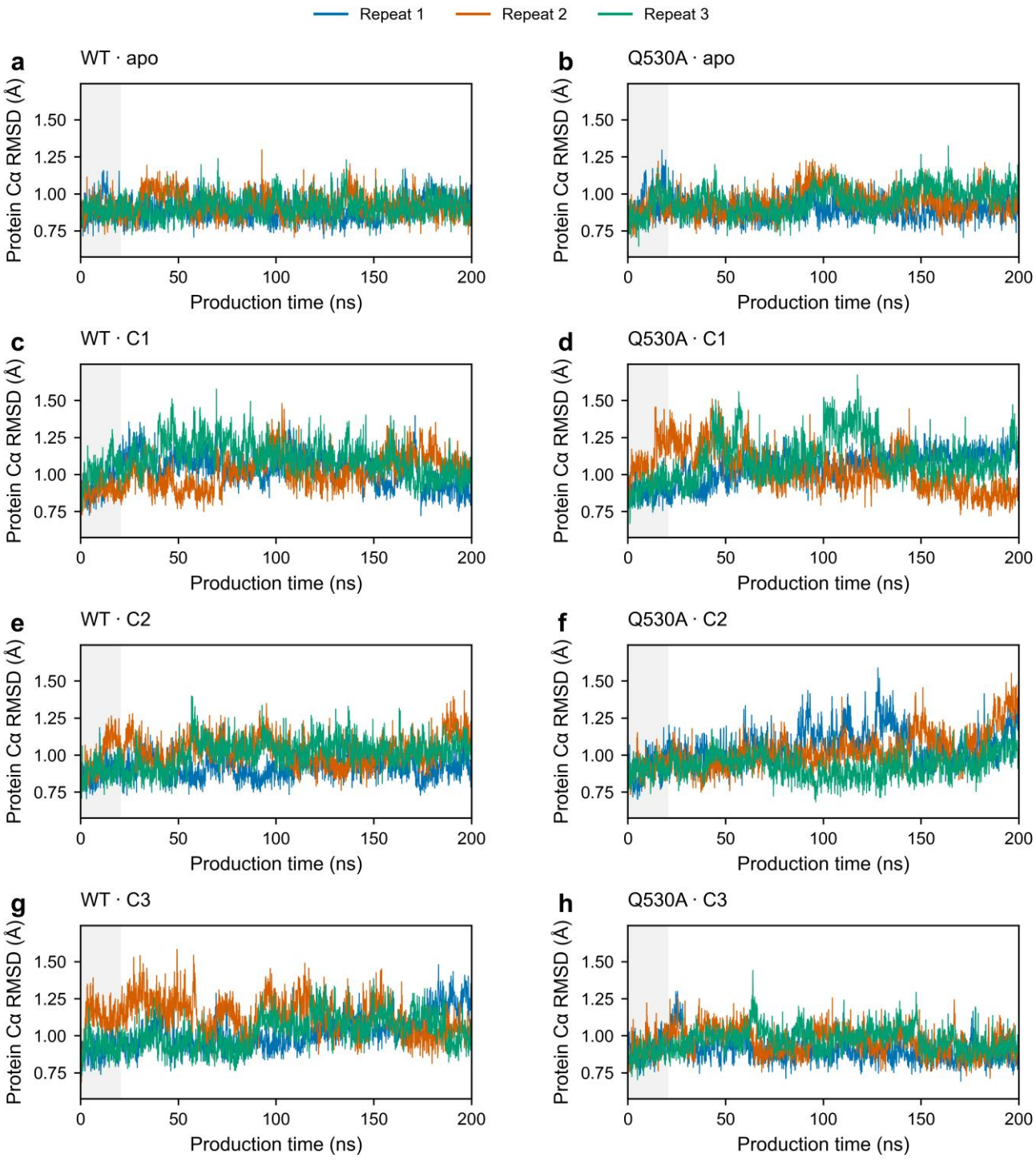

Figure S1. Protein Cα RMSD relative to the prepared WT apo structure in the common receptor frame. Unsmoothed observations begin after preparation. Panels (a,b), (c,d), (e,f) and (g,h) show apo, C1, C2 and C3, respectively; WT is left and Q530A is right. Colors identify the three velocity-seeded trajectories; gray shading marks the excluded initial production interval.

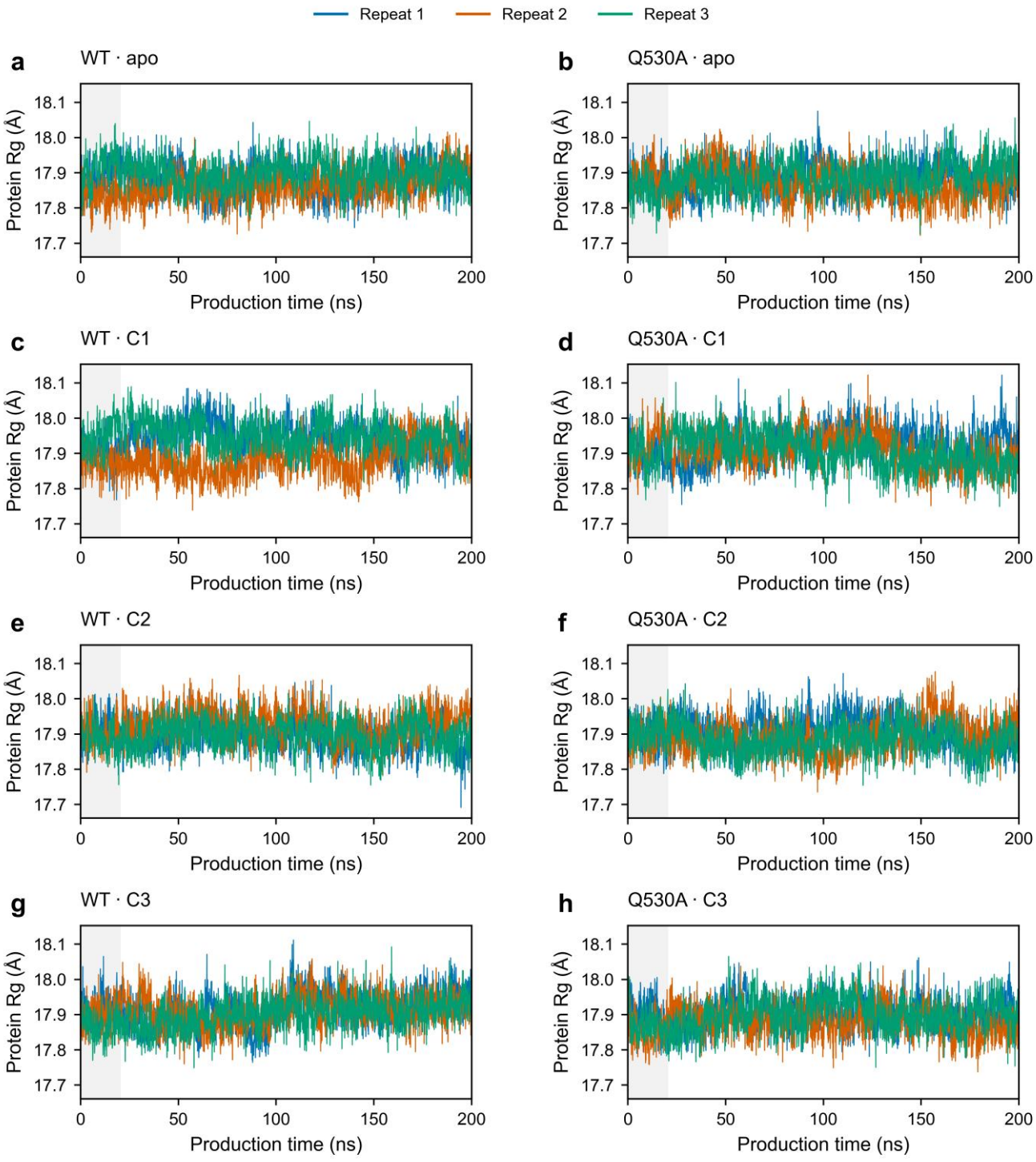

Figure S2. Mass-weighted protein radius of gyration. Unsmoothed traces describe global compactness during production. Panels (a,b), (c,d), (e,f) and (g,h) show apo, C1, C2 and C3, respectively; WT is left and Q530A is right. Colors identify the three velocity-seeded trajectories; gray shading marks the excluded initial production interval.

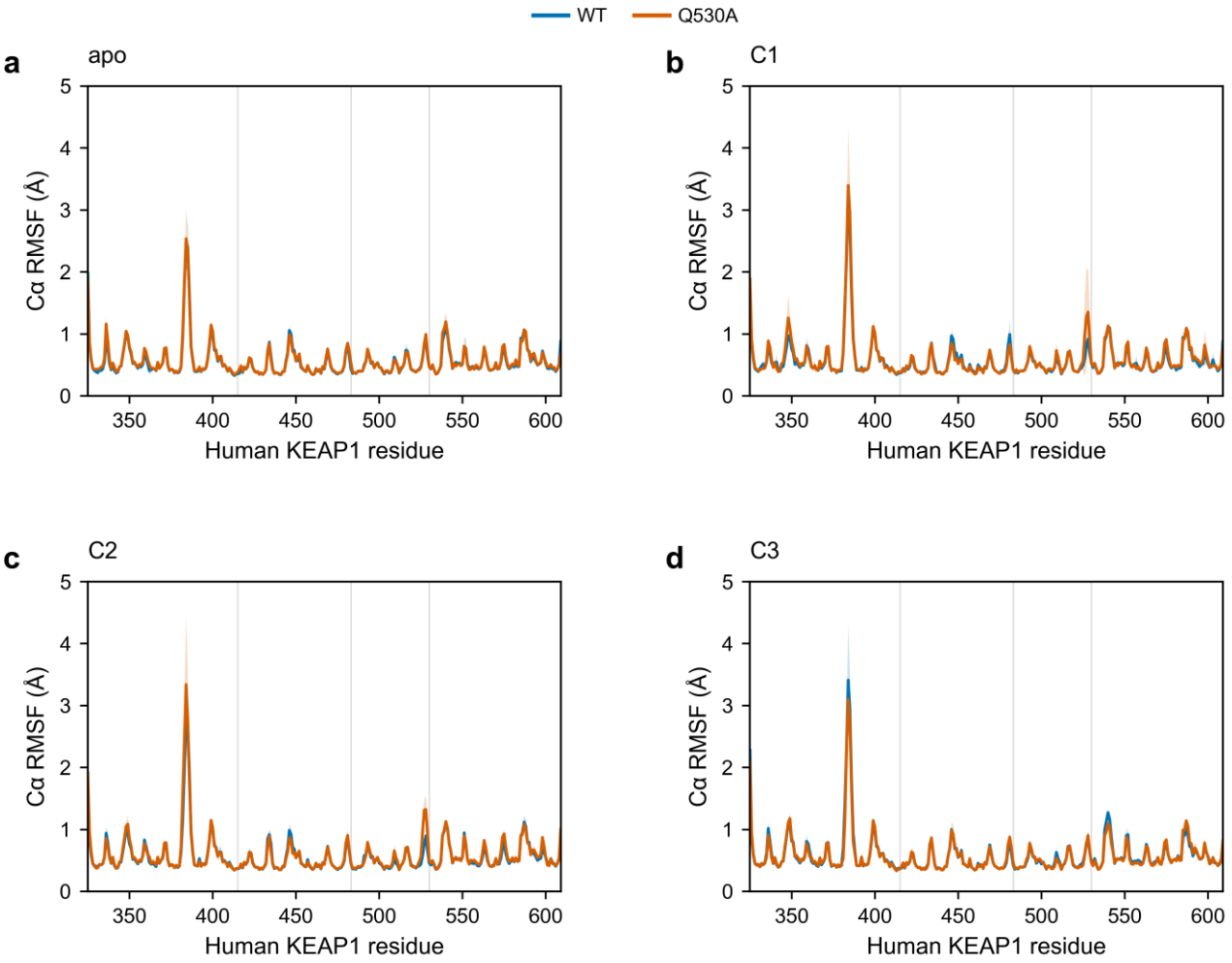

Figure S3. Residue fluctuations in (a) apo, (b) C1, (c) C2 and (d) C3 conditions. Lines and shading show mean Cα RMSF and between-trajectory SD over 20–200 ns. Gray guides mark R415, R483 and residue 530.

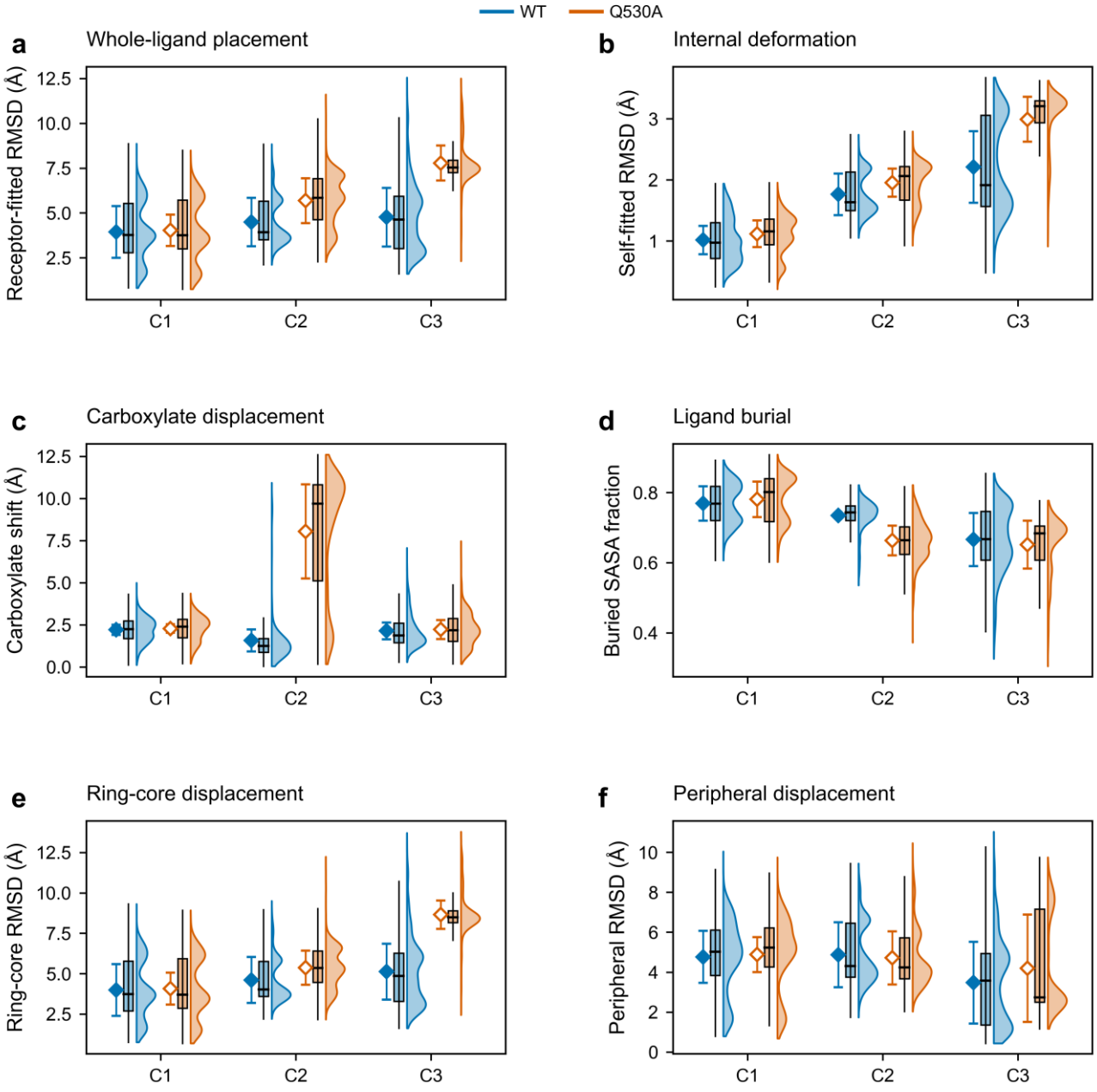

Figure S4. Ligand motion and burial distributions. (a) Receptor-fitted RMSD. (b) Self-fitted RMSD. (c) Carboxylate-centroid displacement. (d) Buried solvent-accessible surface fraction. (e,f) Ring-core and peripheral displacement. Half violins summarize temporal distributions, boxes show median and IQR, and diamonds with error bars show group means and between-trajectory SD.

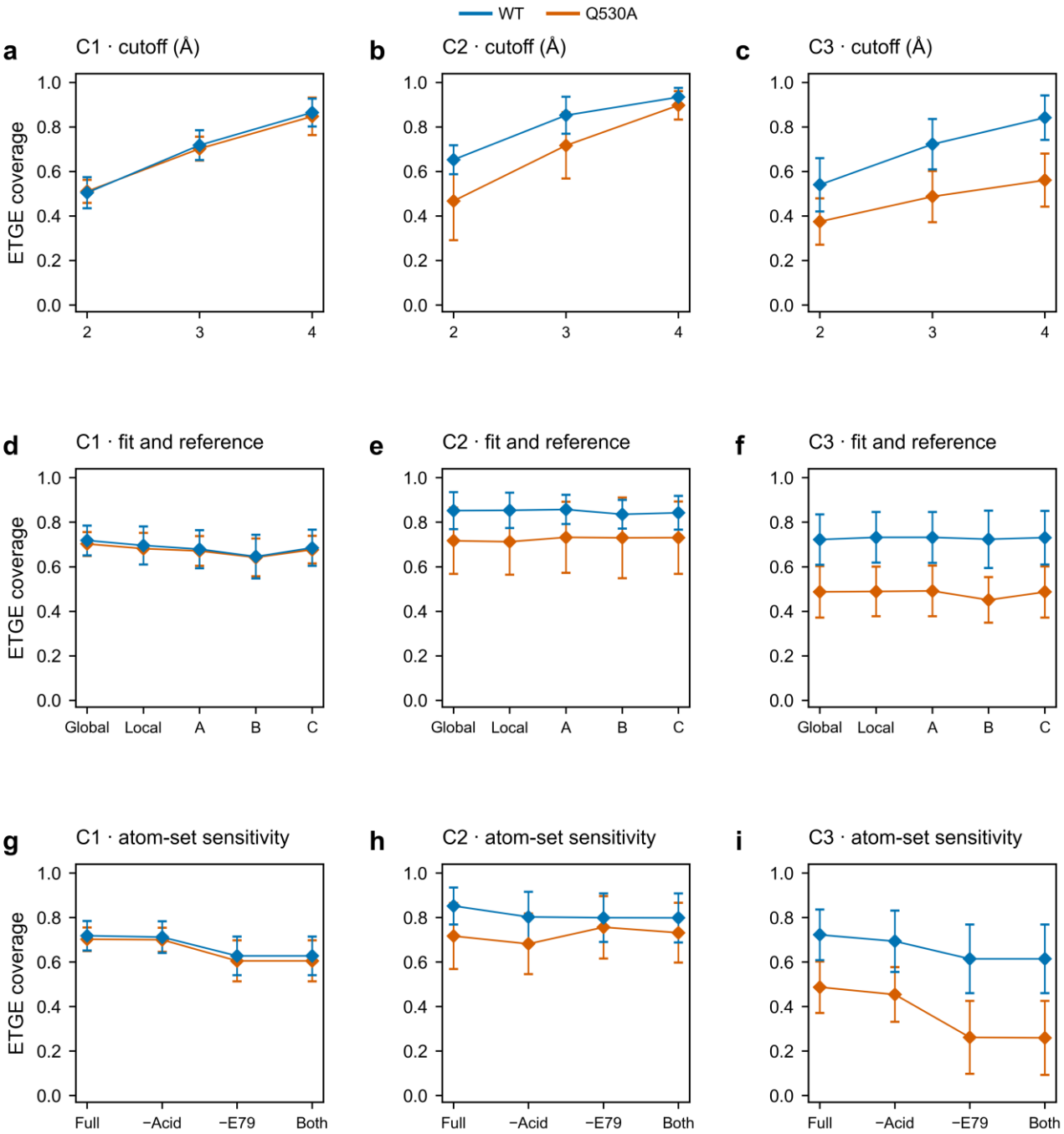

Figure S5. Sensitivity of whole-ETGE reference-site coverage. Columns show C1–C3. (a–c) Distance cutoff. (d–f) Global or local receptor fitting and alternative peptide conformations A–C. (g–i) Descriptor recalculation after excluding ligand carboxylate points (−Acid), reference E79 points (−E79), or both. Symbols and error bars show means and between-trajectory SD.

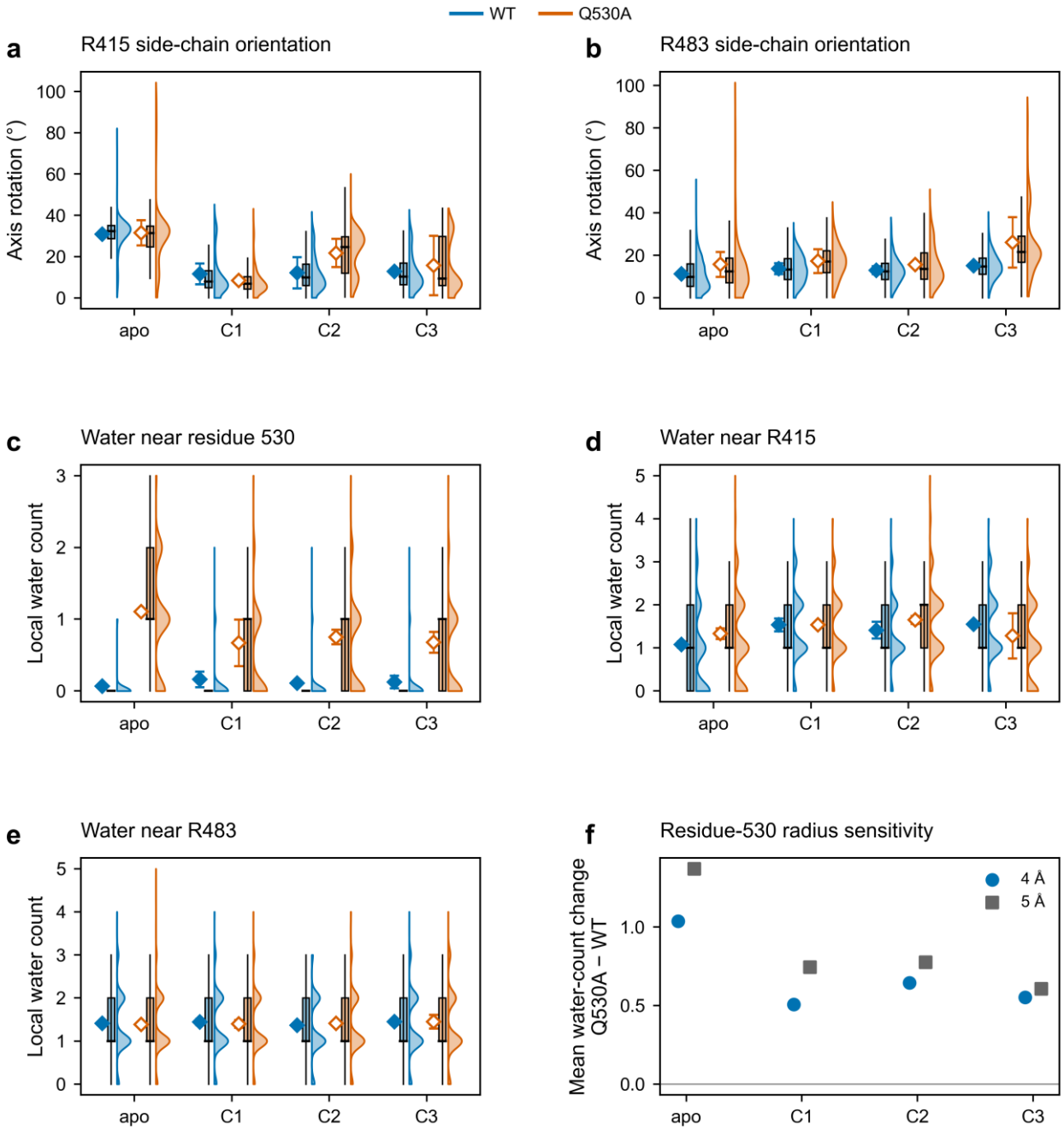

Figure S6. Side-chain geometry and local hydration. (a,b) R415 and R483 Cα–Cζ axis rotation from the prepared WT apo reference. (c) Water oxygen counts within 4 Å of residue-530 Cβ. (d,e) Counts within 3.5 Å of the R415 or R483 guanidinium center. (f) Mean Q530A–WT count changes at 4 and 5 Å around residue 530. Distributions and means use equal trajectory weights; error bars show between-trajectory SD.

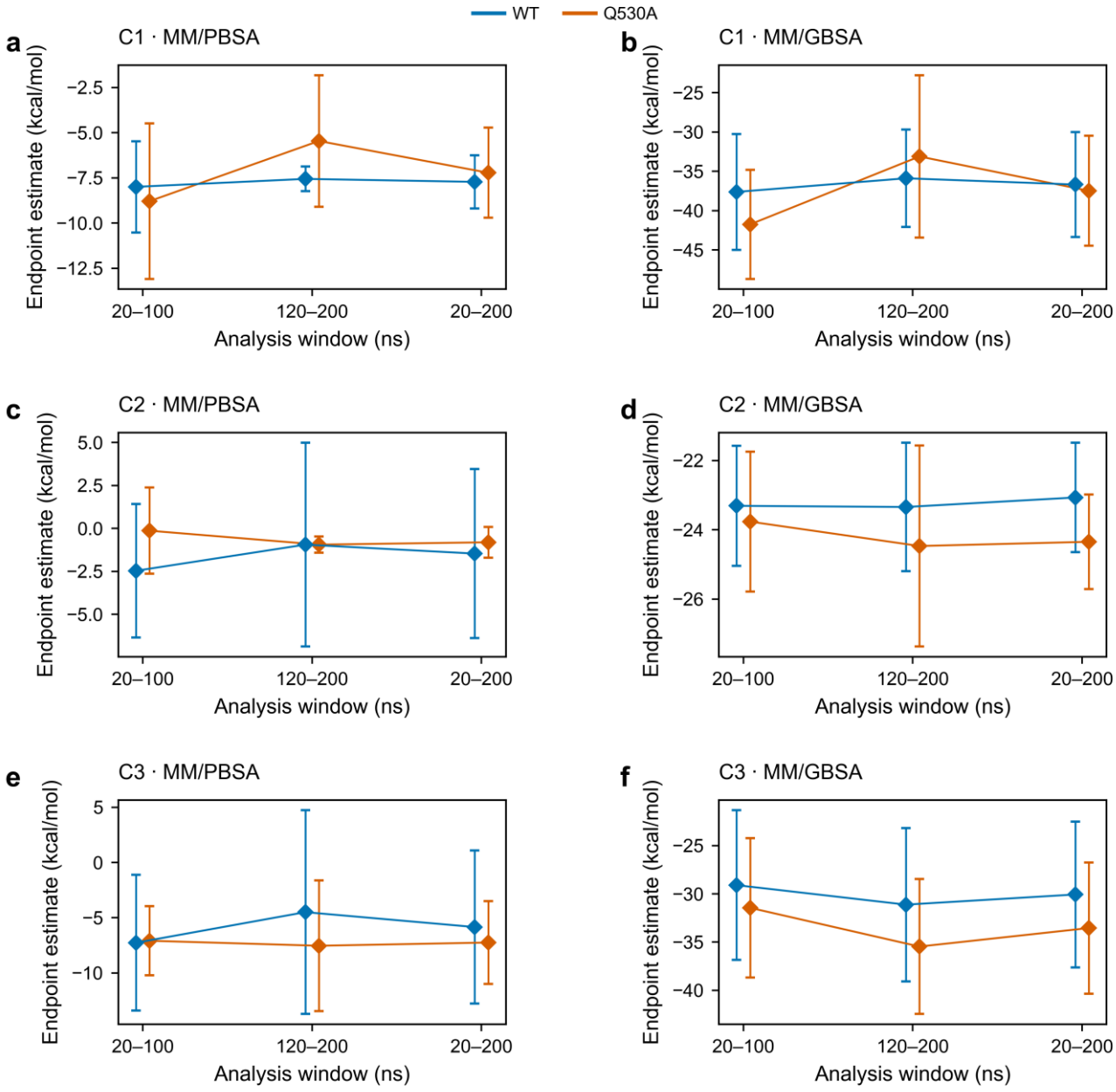

Figure S7. Time-window sensitivity of endpoint energies. (a,b), (c,d) and (e,f) show C1, C2 and C3; MM/PBSA is left and MM/GBSA is right. Symbols and error bars show group means and between-trajectory SD. The equal-duration early and late windows are compared with the complete primary window.

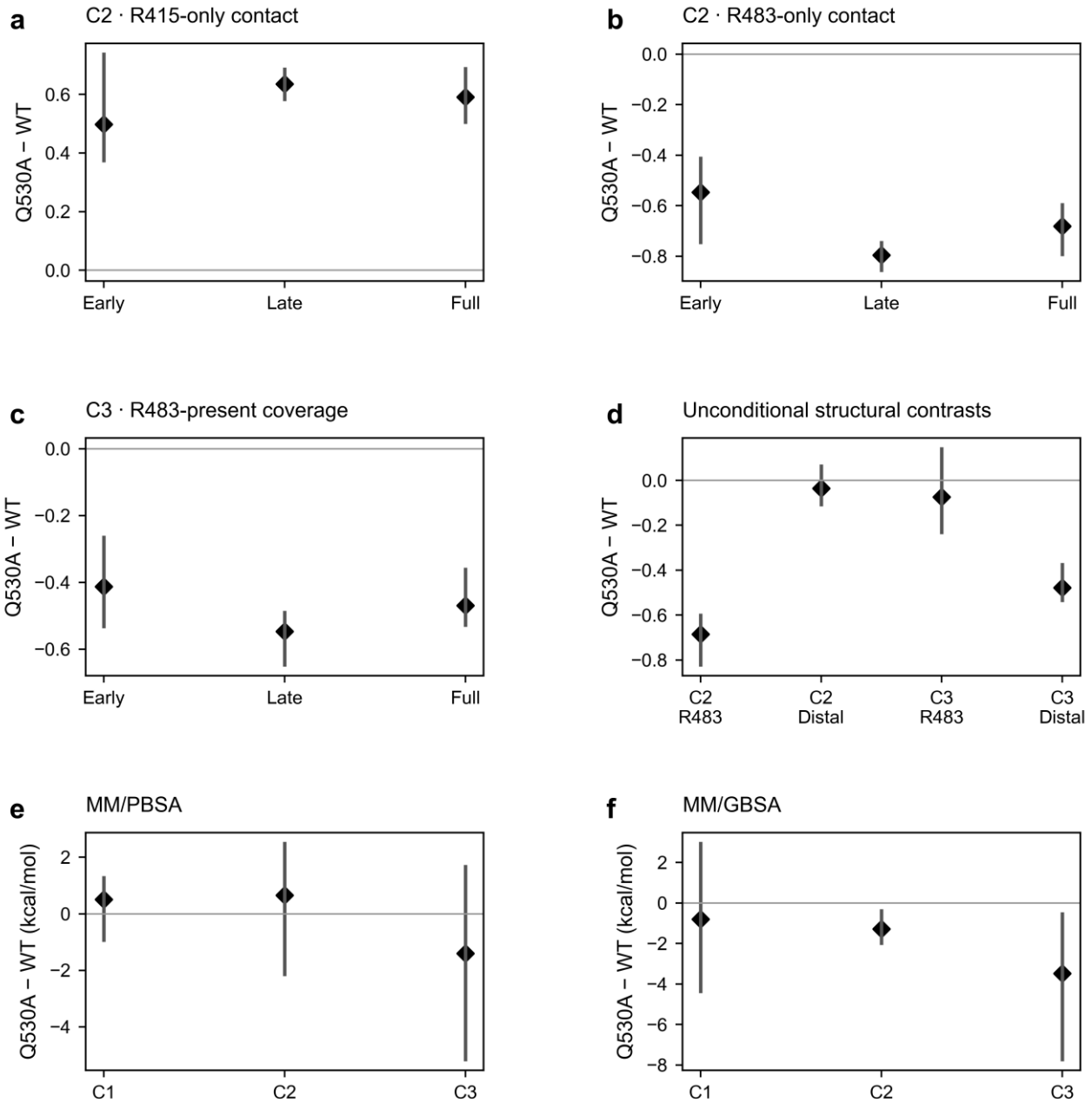

Figure S8. Targeted sensitivity of structural and energetic contrasts. (a,b) C2 exclusive-contact fractions. (c) C3 distal coverage conditional on any R483 hydrogen bond. Early, late and full windows are 20–100, 120–200 and 20–200 ns. (d) Unconditional R483-contact and distal-coverage contrasts. (e,f) MM/PBSA and MM/GBSA contrasts. Diamonds use all trajectories; vertical lines span single-trajectory omissions and are not confidence intervals. The separate stricter R483–carboxylate check and its undefined late-window contrast are described in SM4.

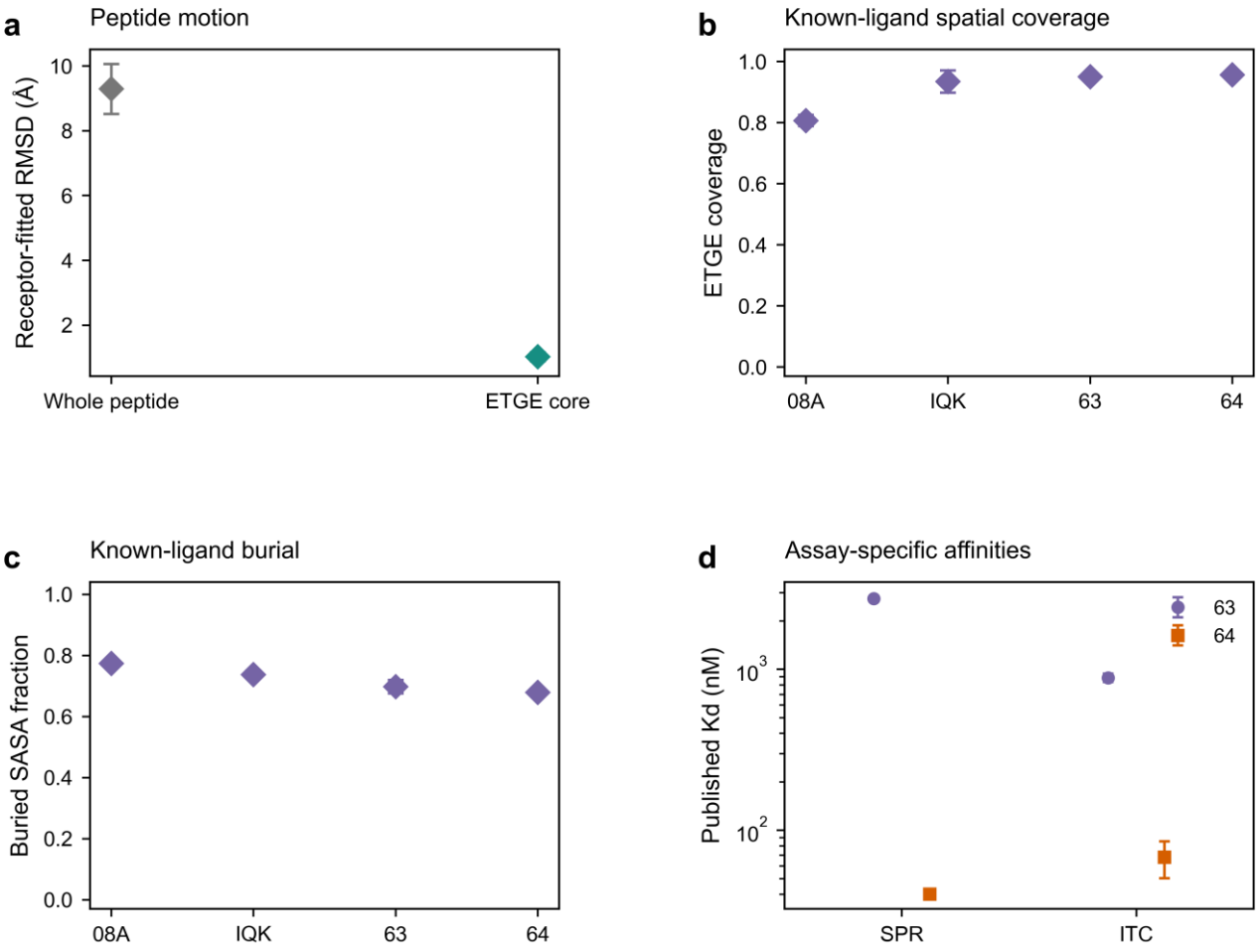

Figure S9. Reference comparisons. (a) Whole-peptide and ETGE-core displacement. (b,c) ETGE-site coverage and ligand burial for known ligands. Structural summaries use the separate 20–100 ns reference cohort and show means with between-trajectory SD. (d) Published affinities of macrocycles 63 and 64, with reported SD, shown separately by assay [13].

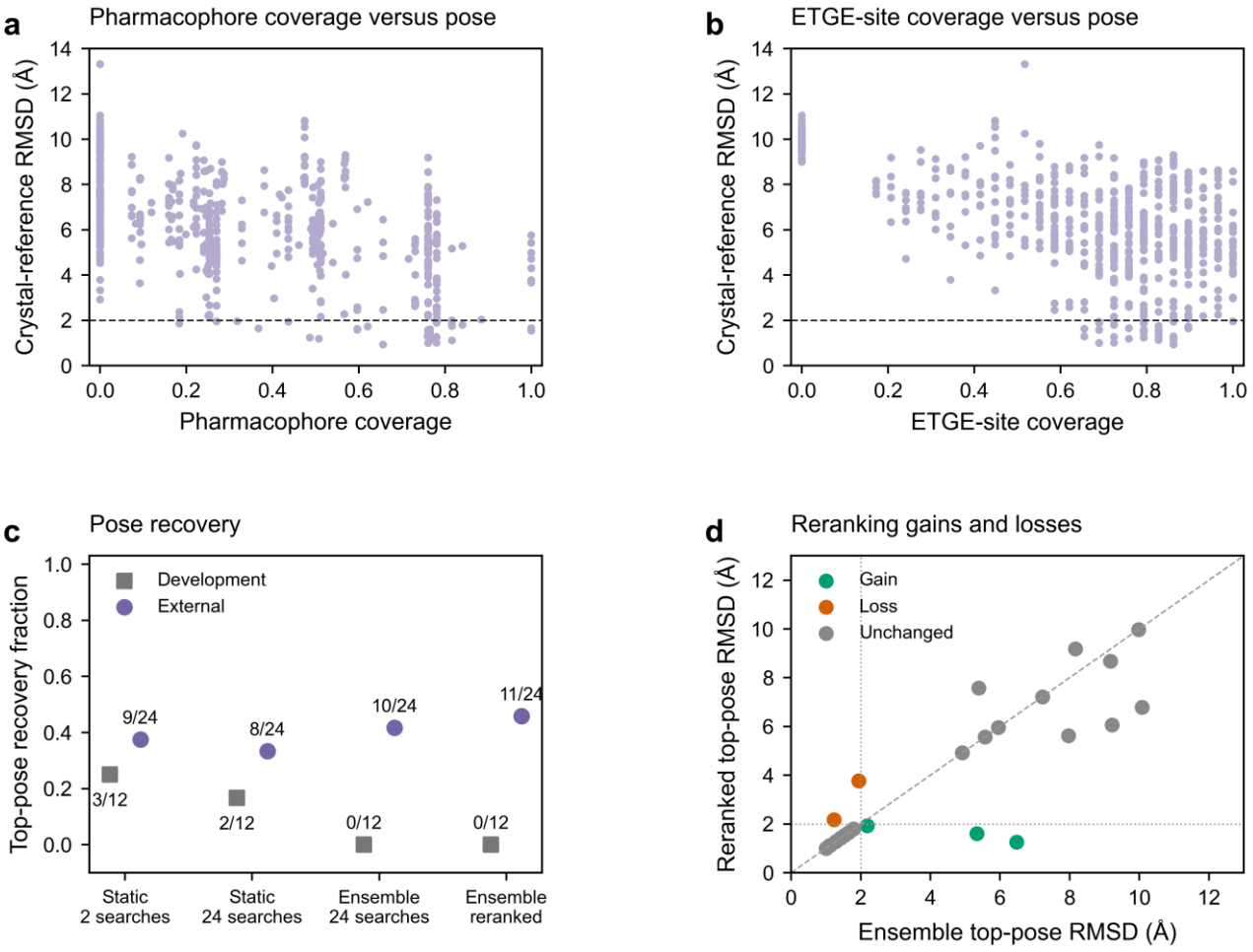

Figure S10. Retrospective pose evaluation. (a,b) Pharmacophore and ETGE-site coverage versus crystallographic-reference RMSD for the complete ensemble pose pool. (c) Top-pose recovery in development and external sets. (d) External top-pose RMSD before and after reranking. The 2 Å guides define pose recovery; poses and search seeds are correlated observations.

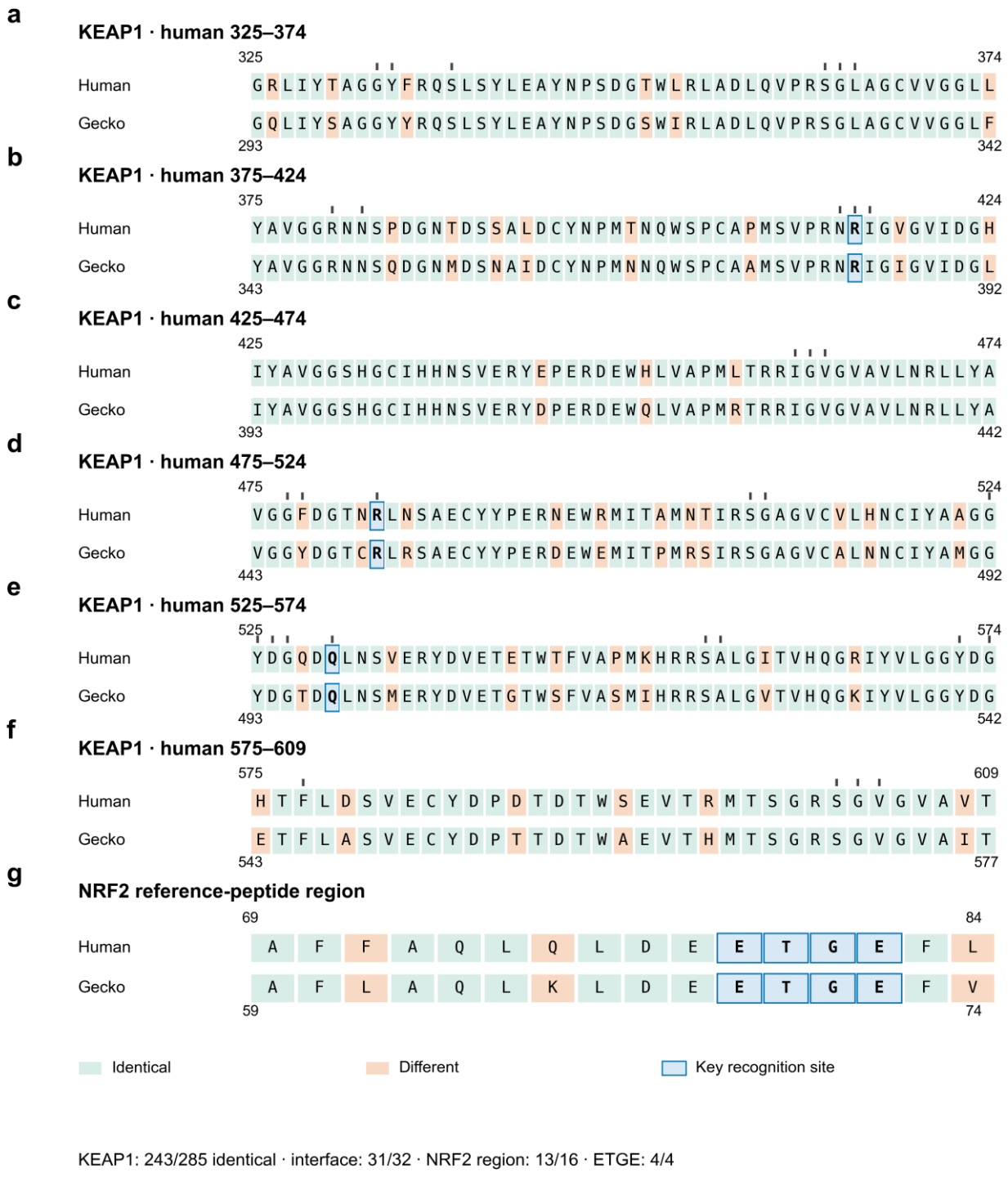

Figure S11. Pairwise human–leopard-gecko sequence alignment. (a–f) The modeled KEAP1 region; dark ticks identify predefined interface residues and blue outlines identify R415, R483 and Q530. (g) The NRF2 reference-peptide region, including ETGE. Green and peach backgrounds indicate identical and differing residues. End labels give species-specific residue numbers. Human F478 corresponds to gecko Y446. Equal-height letters display the two aligned sequences.

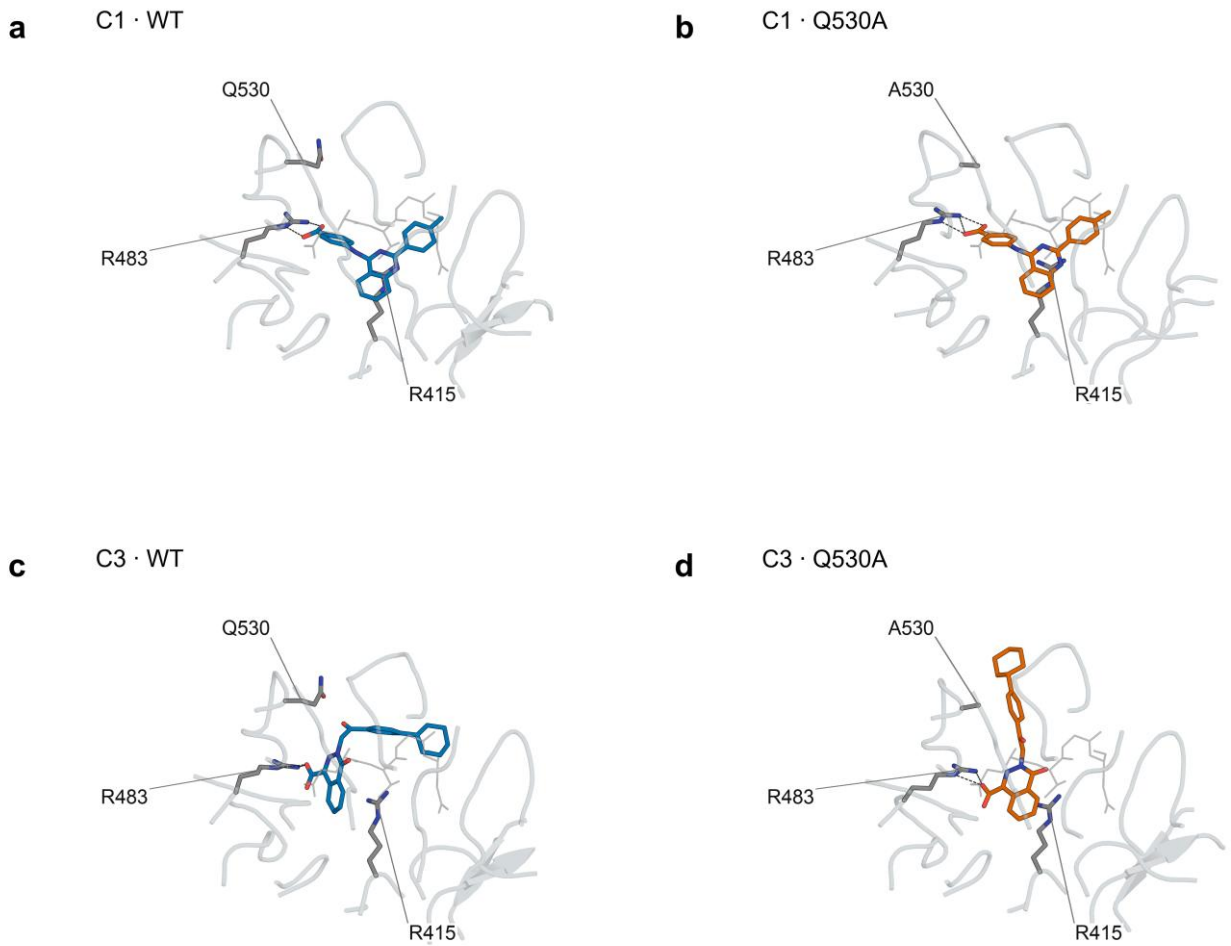

Figure S12. Local contact geometries of C1 and C3. (a,b) WT and Q530A C1. (c,d) WT and Q530A C3. Ligand carbon atoms are blue for WT and orange for Q530A; gray sticks show the ETGE spatial reference, and dashed lines indicate R415/R483 hydrogen bonds.