bioRxiv preprint doi: https://doi.org/10.1101/2022.06.23.497309; this version posted June 23, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

Title: Structural determinants of homophilic adhesion by receptor tyrosine phosphatases

Authors: Iain M. Hay[1,2,3], Maria Shamin[1], Eve R. Caroe[1,4], Ahmed S. A. Mohammed[5], Dmitri I. Svergun[5], Cy M. Jeffries[5], Stephen C. Graham[6], Hayley J. Sharpe[2]* and Janet E. Deane[1]*

Affiliations

[1] Cambridge Institute for Medical Research, University of Cambridge, Cambridge, CB2 0XY, UK.

[2] Signalling Programme, Babraham Institute, Babraham Research Campus, Cambridge CB22 3AT, U.K.

[3] Current address: MRC Laboratory of Molecular Biology, Cambridge CB2 0QH, U.K.

[4] Current address: Retroviral Replication Laboratory, Francis Crick Institute, London NW1 1AT, U.K.

[5] European Molecular Biology Laboratory (EMBL) Hamburg Site, Hamburg, Germany.

[6] Department of Pathology, University of Cambridge, Tennis Court Road, Cambridge CB2 1QP, UK.

* Corresponding authors: Hayley Sharpe (hayley.sharpe@babraham.ac.uk) and Janet Deane (jed55@cam.ac.uk)

Short title: Structural basis of RPTP selectivity

Keywords: cell adhesion, cell contact, RPTP, SAXS

bioRxiv preprint doi: https://doi.org/10.1101/2022.06.23.497309; this version posted June 23, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

ABSTRACT

The type IIB receptor protein tyrosine phosphatases (R2B RPTPs) are cell surface transmembrane proteins that engage in cell adhesion via their extracellular domains (ECDs) and cell signaling via their cytoplasmic phosphatase domains. The ECDs of R2B RPTPs form stable, homophilic, trans interactions between adjacent cell membranes. Previous work has demonstrated how one family member, PTPRM, forms homodimers; however, the determinants of homophilic specificity remain unknown. We have solved the X-ray crystal structure of the membrane-distal, N-terminal domains of PTPRK that form a head-to-tail dimer consistent with intermembrane adhesion. Comparison with the PTPRM structure demonstrates inter-domain conformational differences that may define homophilic specificity. Using small-angle X-ray scattering we determined the solution structures of the full-length ECDs of PTPRM and PTPRK, identifying that both are rigid, extended molecules that differ in their overall long-range conformation. Furthermore, we identify one residue, W351, within the interaction interface that differs between PTPRM and PTPRK and show that mutation to glycine, the equivalent residue in PTPRM, abolishes PTPRK dimer formation *in vitro*. These data support that homophilic specificity within this receptor tyrosine phosphatase family is driven by a combination of shape complementarity and specific but limited sequence differences.

SIGNIFICANCE STATEMENT

Cell-cell contacts are dynamically regulated, in part, by the actions of tyrosine kinases and phosphatases. The R2B family of receptor tyrosine phosphatases combine an adhesive extracellular domain with intracellular catalytic domains that bind and dephosphorylate key cell adhesion and polarity proteins. Previous work demonstrated that the extracellular domains form head-to-tail homodimers but, as the interface was composed of residues conserved across the family, homophilic specificity determinants remained unclear. We have used a range of structural techniques including X-ray crystallography, small angle X-ray scattering and AlphaFold modelling to demonstrate that, despite their similarity, R2B receptors possess significant differences in their overall shape. Our results support that a combination of subtle shape and sequence variations determine homophilic binding.

bioRxiv preprint doi: https://doi.org/10.1101/2022.06.23.497309; this version posted June 23, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

INTRODUCTION

Cell-cell adhesion confers mechanical integrity to tissues that is critical for proper development and barrier function (1). Cell surface adhesion molecules hetero- or homodimerize to connect cells and are linked to the cytoskeleton via intracellular adhesive plaques. These adhesion complexes are highly dynamic and regulated by post translational modifications, in particular tyrosine phosphorylation (2). The actions of protein tyrosine kinases and phosphatases determine phosphotyrosine levels, which can be coupled to cell surface receptor domains, enabling responses to external signals and subsequent regulation of cell adhesion proteins (3,4). The type IIB receptor protein tyrosine phosphatases (R2B RPTPs) combine extracellular adhesion domains with intracellular catalytic phosphatase domains and may act as cell contact sensors in phosphorylation-based signaling events. R2B RPTPs function at cell contact sites and signal through the recruitment and regulation of multiple adhesion plaque protein substrates (5,6).

The four human R2B RPTPs (PTPRK, PTPRM, PTPRT and PTPRU) are type one transmembrane proteins and share a common extracellular domain (ECD) architecture of one MAM (meprin/A5/μ), one immunoglobulin (Ig)-like and four fibronectin type-III (FN) domains, with the most membrane proximal FN domain undergoing furin cleavage in the secretory pathway (7). This is followed by a transmembrane helix, an uncharacterized juxtamembrane domain and two tandem intracellular phosphatase domains (8). Previous studies have shown that PTPRK, PTPRM and PTPRT form only homodimers in cell aggregation assays (9-12). The homophilic (trans) interactions of R2B RPTPs determine their subcellular localization (5,11) and have been proposed to function as spacer clamps between cell membranes (13). Structural and biophysical studies demonstrate that for PTPRM the minimal unit required for dimerization in solution consists of the N-terminal MAM, Ig and first FN domain (hereafter referred to as MIFN1) (14). Dimer formation is pH-dependent, with the ECDs being monomeric at pH 6 and dimeric at pH 8 (14). This is thought to prevent formation of homophilic dimers within the secretory pathway, limiting dimerization to the cell surface at physiological pH (13).

Importantly, although the PTPRM ECD structure shed light on the nature of dimer formation, it did not explain the homophilic binding specificity of R2B receptors.

bioRxiv preprint doi: https://doi.org/10.1101/2022.06.23.497309; this version posted June 23, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

Residues that were identified in the interface and demonstrated as necessary for dimer formation (including Y297, R239, R240 and R409) are conserved across the family (13). An earlier study, using chimeric ECD constructs and clustering assays demonstrated that the MAM domain was necessary but not sufficient for homophilic specificity (15). We therefore set out to determine ECD structures of another family member, PTPRK, to allow a detailed comparison of these dimeric complexes and determine which properties are driving binding specificity.

RESULTS

To investigate dimerization of PTPRK *in vitro*, we expressed and purified both full-length PTPRK-ECD and a construct encoding only the N-terminal MAM-Ig-FN1 domains (PTPRK-MIFN1). One possibility for how homophilic interactions may be maintained is that each family member possesses a unique combination of predicted N-linked glycosylation sites (**Fig. S1**). Previous structural and functional studies with PTPRM used protein that had been deglycosylated or expressed in insect cell-based systems, resulting in non-native glycosylation (13,14). In order to test whether differential glycosylation may play a role in homophilic specificity we expressed PTPRK ECD constructs in the human HEK293-F cell line, with no subsequent deglycosylation, to maintain native glycans. To accurately determine the oligomeric state of these constructs and to test the pH sensitivity of dimerization, we performed size-exclusion chromatography coupled to multi-angle light scattering (SEC-MALS). Based on previous studies with PTPRM, we carried out SEC-MALS analysis at pH 8, where PTPRK would be expected to be dimeric, and pH 6 where monomers were anticipated. For each experiment, protein conjugate analysis was performed to calculate the proportional mass contribution of protein and glycan components. As expected, the full-length PTPRK-ECD was a pH sensitive homodimer, with total protein masses corresponding to dimer and monomer at pH 8 and pH 6, respectively (**Fig. 1A and B**). Similarly, the truncated N-terminal construct PTPRK-MIFN1 showed the same pattern of dimerization (**Fig. 1B and C**), confirming this fragment to be sufficient for formation of the pH sensitive homodimer.

The PTPRK-MIFN1 fragment was crystallized, the X-ray structure solved by molecular replacement using the corresponding fragment of the PTPRM-ECD (PDB ID: 2V5Y, (13)), and the structure refined to 3.0 Å resolution (**Fig. 1D** and **Table S1**). The

bioRxiv preprint doi: https://doi.org/10.1101/2022.06.23.497309; this version posted June 23, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

structure comprises a MAM domain (residues 31-192) that forms an extensive intramolecular interface with the Ig domain (residues 193-290), followed by the FN1 domain (residues 291-388) that makes few inter-domain contacts with the Ig domain (discussed below). Three glycosylation sites were observed in the electron density maps at asparagine residues N101 and N140 in the MAM domain and N211 in the Ig domain. A surface loop on the Ig domain encompassing residues 221-225, which possesses low conservation across the R2B family (**Fig. S1**), was unable to be built due to poor quality electron density. Two chains of PTPRK-MIFN1 were present in the asymmetric unit of the crystal structure forming a head-to-tail dimer and burying a surface area of ~1750 $Å^2$ (**Fig. 1E**). The dimer interface is formed between the MAM-Ig domain from one molecule and the FN1 domain of the opposing molecule, similar to that observed previously for PTPRM (13). Of the three observed glycosylation sites in PTPRK-MIFN1, N140 and N211 are located on the surface distal from the dimer interface and therefore are not likely to be important for determining homophilic binding specificity (**Fig. 1D and E**). N101 is located at the edge of the interface and may play a role in stabilization of the interaction.

To probe what might be contributing to binding specificity, the dimer interfaces were analyzed in further detail. The PTPRK-MIFN1 dimer is stabilized by several hydrogen bonds and salt bridges present at the MAM-Ig:FN1 interface (**Fig. 2A**). To aid understanding of these interactions the residue number will be accompanied by the domain they lie within, for example N207 in the Ig domain will be $N207^{Ig}$. A loop encompassing residues 348-354 at the C-terminal end of the FN1 domain forms several key interactions. The side chain of $K349^{FN1}$ forms hydrogen bonds with the opposing $N101^{MAM}$, as well as the $W160^{MAM}$ backbone carbonyl oxygen (**Fig. 2B**). Notably, the orientation of the $W160^{MAM}$ carbonyl oxygen is caused by a *cis*-peptide bond with the adjacent $P161^{MAM}$, a conformation that is stabilized by the side chain of $W160^{MAM}$ stacking with the N-acetylglucosamine of the $N101^{MAM}$ glycan. $W351^{FN1}$ and $H352^{FN1}$ both occupy a deep groove located at the MAM-Ig boundary of the opposing molecule, burying ~90% of their solvent accessible surface area. $W351^{FN1}$ lies in a hydrophobic pocket created by the aliphatic portion of the $R200^{Ig}$ side chain and $F159^{MAM}$. $R200^{Ig}$ additionally forms a hydrogen bond with the backbone carbonyl oxygen of $K349^{FN1}$. $H352^{FN1}$ occupies a hydrophilic pocket, forming hydrogen bonds with the side chains of $T103^{MAM}$ and $S157^{MAM}$, and the $H352^{FN1}$ side chain imidazole

bioRxiv preprint doi: https://doi.org/10.1101/2022.06.23.497309; this version posted June 23, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

ring stacks against the side chain of H197^{MAM}. Finally, D354^{FN1} forms a salt bridge with R250^{Ig} (**Fig. 2B**).

All but one of these key residues are completely conserved across the R2B receptor ECDs, the difference being that PTPRK W351^{FN1} is a glycine in PTPRM (**Fig. S1**). Despite lacking a bulky side chain at the equivalent position of W351, the binding groove between F159^{MAM} and R200^{Ig} that is occupied by this residue is maintained in PTPRM (**Fig. 2C**). To investigate the role of residue W351^{FN1} in dimer formation, we expressed and purified a W315G mutant form of PTPRK-MIFN1 for SEC-MALS analysis to test pH-dependent homodimer formation. Mutation of W351 in PTPRK to glycine was sufficient to inhibit dimerization, remaining primarily monomeric at pH 8 (**Fig. 2D**). Unfortunately, the reciprocal mutation G341W in the PTPRM-MIFN1 resulted in protein that aggregated to non-discrete oligomers in solution and was not suitable for further analysis. While these experiments demonstrate that W351 is required for PTPRK dimerization *in vitro*, they don't explain the basis of homotypic selectivity as PTPRM has a surface cleft that could accommodate the W315 side chain.

Detailed comparison of the PTPRK-MIFN1 structure with the equivalent region of PTPRM reveals a significant hinge movement (approximately 15°) of the FN1 domain relative to the MAM-Ig domains (**Fig. 3A**). This hinge movement is sufficient to induce a steric clash in the dimer interface when the MAM-Ig domains of PTPRK and PTPRM are superposed (**Fig. 3B**). This inter-domain movement may represent inherent flexibility between these domains or may represent rigid structural differences that contribute to homophilic specificity. For elongated multi-domain proteins, such as the R2B family, even small changes in inter-domain orientation have the potential to result in large conformational differences over the length of the full ECD. Superposing the PTPRK-MIFN1 structure on the PTPRM-ECD structure (encompassing the MAM-Ig-FN1-3 domains, PDB ID: 2V5Y), using the MAM-Ig domains for alignment, illustrates these potential differences (**Fig. 3C**). Models of the full-length ECDs of PTPRM and PTPRK generated using AlphaFold2 (AF2) possess significantly different inter-domain orientations relative to each other and also to the available crystal structures (**Fig. S2**). This lack of confidence in the inter-domain positions is quantified in the predicted

bioRxiv preprint doi: https://doi.org/10.1101/2022.06.23.497309; this version posted June 23, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

alignment error (PAE) plots for these AF2 models (**Fig. S2B**). Despite this conformational range, the hinge movement observed in the PTPRK-MIFN1 structure determined here is not captured by the AF2 PTPRK models, suggesting the full-length ECDs may be even more conformationally distinct than is predicted by these models (**Fig. S2C**). These observations support the need for experimental approaches to determine the long-range shape of the ECD domains.

The available X-ray structure of the PTPRM ECD (PDB ID: 2V5Y, (13)) is missing the final FN4 domain, limiting our understanding of the relative orientation of this domain. We therefore determined the crystal structure of the PTPRM-FN3-4 domains (**Fig. 3D** and **Table S1**). Crystals grew in two space groups, $P2_12_12_1$ and $P3_221$, possessing two and four molecules per asymmetric unit, respectively (**Fig. S3A and B**). All six chains are essentially identical in structure (pairwise RMSDs over 220 Cα atoms range from 0.44 to 0.92 Å, **Fig. S3C**). The main difference between the protein structures is that one chain in the $P2_12_12_1$ crystal has a disordered loop (residues 628-641) that in all other chains was ordered via intermolecular interactions (**Fig. S3A and B**). This loop is anchored at one end via a disulfide bond within the FN4 domain (C642 to C716) and the sequence encompasses the furin-cleavage site of PTPRM, illustrating that this site is surface accessible within a conformationally dynamic loop but that cleavage would not significantly destabilize the fold of the FN4 domain (**Fig. 2D**). There are two predicted glycosylation sites in FN3-4, one in each domain: N544 in FN3; and N598 in FN4. In both crystal forms one chain is missing the glycan on N598, indicating that PTPRM-FN4 not only adopts conformational differences in the furin-cleavage loop but can also undergo heterogeneous glycosylation.

Although the previous structure of the PTPRM ECD had no interpretable electron density for the FN4 domain, it was present in the construct. Superposition of the FN3 domains from our structure with this previous PTPRM structure (13) demonstrates that the FN4 domain can be accommodated without significant steric clashes with surrounding chains, and FN4 co-localizes with weak and disconnected electron density that is unmodeled in the deposited structure. This is consistent with the FN3-4 domains adopting similar conformations when crystallized in isolation or as part of the full-length ECD. This alignment therefore allowed the generation of a full-length ECD chimeric structure based on crystallographic data (**Fig. 3E**). Superposition of this

bioRxiv preprint doi: https://doi.org/10.1101/2022.06.23.497309; this version posted June 23, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

chimeric PTPRM structure and the full-length PTPRM-ECD AF2 models (aligned using the MAM-Ig domains) reveals the extent to which these models for the exact same protein can diverge (**Fig. 3E**). This highlights the difficulty in confidently determining long-distance conformational differences between PTPRM and PTPRK ECDs from crystallographic snapshots or *ab initio* models alone.

To probe the overall shape of these domains in solution we implemented small-angle X-ray scattering coupled to SEC and MALS (SEC-SAXS-MALS). The structural parameters determined for purified PTPRK and PTPRM ECDs confirm that both proteins elute as glycosylated monomers at pH 6 (**Table S2**). The experimental SAXS profiles for PTPRK-ECD (**Fig. 4A**) and PTPRM-ECD (**Fig. 4B**) appear quite similar, indicating conserved global structural features in both proteins. Each of the ECDs are extended particles and both display highly skewed real-space distance distributions ($p(r)$ profiles) with a maximum particle dimension, D_{max}, of 26 nm (**Fig. 4C**). However, the $p(r)$ profiles at shorter distances are not identical, confirming differences in the local domain organization of these proteins. An overlay of the scaled SAXS data (**Fig. S4A**) also shows deviations between the PTPRK and PTPRM scattering intensities at lower angles, in particular spanning an s-range of 0.5–1.6 nm^{-1}, which corresponds to real-space distances of 3.9–12.5 nm. These observations, combined with a consistent 0.2 nm increase in the radius of gyration (R_g) and 0.1 nm decrease in the R_g of cross-section (R_g^c) of PTPRM compared to PTPRK (**Table S2; Fig. S4B**), point to small but detectable differences between the spatial disposition of the domains in the two constructs. With respect to the overall structural sampling of the ECDs, the dimensionless Kratky plots of the SAXS data (**Fig. S4C**) display a monotonic increase for $sR_g < 5$, typical of particles with extended but relatively rigid conformations and not of proteins with domains connected by flexible linkers sampling diverse states in solution (16).

A*b initio* shape analysis with the program DAMMIN (17) yielded narrow (~3–4 nm), extended and slightly bent dummy atom models for both ECDs (**Fig. 4D and E**). Pseudo-atomic models of the PTPRK and PTPRM ECDs were generated by combining the available crystal structures with AF2 predictions for uncharacterized domains, with glycans added to the relevant residues. Initial hybrid models for the PTPRK and PTPRM ECDs fit poorly to the SAXS data, with discrepancy $\chi^2 = 1.78$ and

bioRxiv preprint doi: https://doi.org/10.1101/2022.06.23.497309; this version posted June 23, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

probability of systematic deviations between the model fit and the scattering data (CorMap *P,* (18)) of 5.0×10^{-6} for PTPRK and χ^2 = 1.42 (CorMap $P = 1.0\times10^{-8}$) for PTPRM. Rigid body refinement of these initial models to the SAXS data using CORAL (19), which allowed repositioning of the MAM-Ig and FN domains with respect to each other, yielded pseudo-atomic models with significantly improved fits, χ^2 = 1.17 (CorMap P = 0.072) for PTPRK-ECD and χ^2 = 1.10 (CorMap P = 0.293) for PTPRM-ECD (**Fig. 4D and E, Table S2**). Superposition of these final pseudo-atomic models (**Fig. 4F**) confirms that both adopt extended conformations in solution, but that PTPRK appears more 'bent/twisted' compared to PTPRM. Such conformational differences are consistent with the observed differences in the $p(r)$ profiles for PTPRK and PTPRM, and with the smaller radius of gyration (R_g) and larger R_g cross-section (R_g^c) obtained from the SAXS data for PTPRK. Furthermore, the conformations of the PTPRM and PTPRK ECDs differ in shape sufficiently that the SAXS-based hybrid model of one construct does not fit the SAXS data for the other (PTPRK-ECD, $\chi^2_{\text{PTPRM-SAXS}}$ = 1.75, CorMap $P = 9.8\times10^{-12}$; PTPRM-ECD, $\chi^2_{\text{PTPRK-SAXS}}$ = 1.26, CorMap $P = 3.6\times10^{-5}$). These SAXS models indicate that the overall shape of the R2B ECDs differ from each other, which may drive homophilic specificity.

To probe what might be contributing to the different shapes of PTPRK and PTPRM ECDs, the X-ray structures were analyzed in closer detail with particular focus on the intramolecular contacts between the MAM-Ig and FN1 domains. Comparative analysis of PTPRK and PTPRM shows a large degree of conservation at the intramolecular interface between the Ig and FN1 domains (**Fig. S5**). In PTPRK, the side chain of $N207^{Ig}$ forms hydrogen bonds with the side chain of $R367^{FN1}$ and the backbone carbonyl oxygens of $P368^{FN1}$ and $E370^{FN1}$. The former two bonds are conserved in PTPRM and inspection of electron density maps from PDBe (20,21) supports the third also being conserved. The PTPRK $D320^{FN1}$ side chain and backbone amide nitrogen form hydrogen bonds with the backbone amide nitrogen and carbonyl oxygen, respectively, of $E290^{Ig}$. Inspection of PTPRM electron density maps also supports this interaction. One subtle difference is that in PTPRM $E195^{Ig}$ forms a salt bridge with $R357^{FN1}$, whereas in PTPRK the equivalent $E205^{Ig}$ residue instead forms an intradomain salt bridge with $R289^{FN1}$, which is a lysine residue in PTPRM. Furthermore, on a loop of the PTPRK Ig domain, near the FN1 interface, there is an

bioRxiv preprint doi: https://doi.org/10.1101/2022.06.23.497309; this version posted June 23, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

N-linked glycan at N211 that is absent in PTPRM (**Fig. 3A**). This glycan sits near the apex of the hinge movement of FN1 and lies on the opposite side of PTPRK from the dimer interface. Although the modelled glycan does not make any stable contacts with the FN1 domain, the presence of a bulky glycan tree here could cause the FN1 domain to 'swing' away from the Ig domain. Similar steric hindrance based conformation changes have been observed previously (22,23), for example, in the glycan-mediated regulation of plasminogen activator inhibitor-1 (24).

DISCUSSION

The ECDs of R2B RPTPs engage in homophilic (trans) interactions to bridge membrane contact sites. Previous studies using PTPRM identified how these ECDs form head-to-tail dimers but did not investigate how homophilic specificity is determined. Here we describe structures of a related family member, PTPRK, using X-ray crystallography and SAXS. The crystal structure of the N-terminal MAM-Ig-FN1 construct allowed direct comparison with the PTPRM structure, showing that PTPRK can engage in a similar head-to-tail dimer, while also revealing differences in the relative orientations of the N-terminal MAM-Ig and FN1 domains that mediate dimerization. Detailed analysis of these interdomain contacts in PTPRK versus PTPRM suggests that the hinge movement between the MAM-Ig and FN domains is not driven by substantial sequence differences, but is instead driven by a small number of subtle differences. These include the presence of a single mostly-disordered N-linked glycan and possibly a conservative amino acid change (R to K) at the domain interface. These small differences position the PTPRK FN1 domain such that it is likely to be structurally incompatible with forming a heterodimer with PTPRM. Importantly, SAXS analysis confirms that the full-length ECDs of PTPRM and PTPRK adopt differing rigid, extended conformations in solution, with PTPRK being more bent/twisted.

Detailed analysis of the crystallographic PTPRK and PTPRM homodimers identifies that the vast majority of residues involved in the dimerization interface are conserved across the R2B family of proteins, indicating that any contributions of primary sequence composition to determining homophilic binding specificity are likely to be subtle. Indeed, previous work identifying the sequence determinants of dimer formation using cell clustering assays mutated residues that are conserved across the

bioRxiv preprint doi: https://doi.org/10.1101/2022.06.23.497309; this version posted June 23, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

R2B family, supporting that the crystallographic dimer is the relevant complex but not how homophilic specificity is established. The pH-dependency of dimer formation shown here for PTPRK is the same as that for PTPRM, supporting that both proteins possess similar electrostatic complementarity. We identify one amino acid difference in the interaction interface (W351 in PTPRK is a glycine in PTPRM) and that introduction of this single mutation is sufficient to inhibit PTPRK homodimerization.

These data support that subtle differences in sequence, N-linked glycosylation and shape combine to drive the specific homophilic intercellular interactions of the R2B RPTP family of proteins. These observations are reminiscent of the binding determinants for type I cadherins (25). These proteins possess tandem cadherin domains where subtle residue changes in the interaction interface, combined with distinct long-axis angles, determine hetero- and homodimer affinities. Finally, the structural observations made here have been determined with isolated ECDs separate from their associated membrane domains. How these proteins are oriented on the cell surface is highly challenging to measure but may further contribute to homophilic specificity of R2B family members in the context of a membrane interface.

EXPERIMENTAL PROCEDURES

Plasmids and constructs

Amino acid numbering for DNA constructs is based on the following sequences; human PTPRK (Uniprot Q15262-3) and human PTPRM (Uniprot P28827-1). For recombinant expression in mammalian cells, PTPRK full ECD (residues 28-752) and MIFN1 (residues 28-385) and PTPRM full ECD (residues 19-742) and FN3-4 (residues 478-723) were PCR amplified and cloned into the pHLSec vector using AgeI and KpnI restriction sites allowing in-frame expression with an N-terminal secretion signal and a C-terminal Lys-His_6 tag.

Protein expression and purification

Expression of recombinant extracellular proteins was performed by transient transfection in HEK-293F cells using polyethylenimine (PEI, Sigma Aldrich). 300 mL of culture at a density of 1.0 x 10^6 cells/mL was transfected with 300 µg DNA using 450 µg PEI. Culture medium was harvested 72 h post-transfection and secreted His-

bioRxiv preprint doi: https://doi.org/10.1101/2022.06.23.497309; this version posted June 23, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

tagged proteins purified using Ni-NTA agarose (Qiagen) in batch mode. Beads were packed into a 20 mL gravity column, washed with 20 mL of wash buffer (50 mM Tris-HCl, pH 8, 500 mM NaCl, 20 mM imidazole) and eluted using wash buffer containing 250 mM imidazole. Eluted protein was further purified by size exclusion chromatography (SEC). For PTPRK and PTPRM ECD constructs SEC was performed using a HiLoad Superdex 200 pg 16/600 column (Cytiva) equilibrated in 100 mM Tris-HCl, pH 8, 250 mM NaCl. For the PTPRM FN3-4 constructs SEC was run on a HiLoad Superdex 75 pg 16/600 column (Cytiva) equilibrated in 20 mM Tris-HCl, pH 7.4, 150 mM NaCl.

Multi-angle Light Scattering (MALS)

MALS experiments were performed immediately following SEC (SEC-MALS) by inline measurement of static light scattering (DAWN 8+; Wyatt Technology), differential refractive index (Optilab T-rEX; Wyatt Technology), and UV absorbance (1260 UV; Agilent Technologies). Samples (100 µl) at 1 mg/mL were injected on to a Superdex 200 Increase 10/300 GL column (Cytiva) equilibrated in pH 8 purification buffer (100 mM Tris-HCl, pH 8, 250 mM NaCl) at a flow rate of 0.4 mL/min. The molar masses of the major SEC elution peaks were calculated in ASTRA 6 (Wyatt Technology) using a protein dn/dc value of 0.185 mL/g. For determination of protein and glycan fractions, conjugate analysis was performed in ASTRA 6, using a glycan (modifier) dn/dc = 0.14 mL/g and theoretical UV extinction co-efficients calculated using ProtParam (26). For experiments at pH 6, ECD peak fractions were collected from pH 8 experiments and buffer exchanged using centrifugal concentration units (Amicon, Merck) into pH 6 purification buffer (50 mM MES, pH 6, 250 mM NaCl). SEC-MALS experiments at pH 6 were then carried out as for pH 8 experiments.

Crystallization

Crystallization experiments were carried out in 96-well nanolitre-scale sitting drops (200 nL of purified protein with 200 nL of precipitant) equilibrated at 20°C against 80 µL reservoirs of precipitant. PTPRK-MIFN1 crystallization was attempted using 9.5 mg/mL protein and diffraction quality crystals grew against a reservoir of 100 mM HEPES, pH 7, 10% (w/v) polyethylene glycol (PEG) 6000. Two different PTPRM FN3-4 crystals grew in related conditions. The $P2_12_12_1$ crystals grew following equilibration of purified protein at 11.2 mg/mL against a reservoir containing 200 mM ammonium

bioRxiv preprint doi: https://doi.org/10.1101/2022.06.23.497309; this version posted June 23, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

nitrate and 20% (w/v) PEG 3350. The $P3_221$ crystals were grown following microseeding as described previously (27) using a seed stock made from the $P2_12_12_1$ crystals combined with purified protein at 9.1 mg/mL against a reservoir containing 100 mM ammonium nitrate and 10% PEG 3350. All crystals were cryoprotected in reservoir solution supplemented with 20% (v/v) glycerol and flash-cooled by plunging into liquid nitrogen.

X-ray data collection and structure solution

X-ray diffraction datasets were recorded at Diamond Light Source (DLS) on beamlines I03, I04 and I04-1 (**Table S1**). Diffraction datasets were indexed and integrated using the XIA2 DIALS pipeline (28,29). For PTPRK-MIFN1, the initial structure was solved by molecular replacement (MR) using Phaser (30), with the MAM, Ig-like and first FN domains of the human PTPRM full ECD (PDB ID: 2V5Y, (13)) as a search model. For PTPRM-FN3-4, there was no model available (at the time) that encompassed both domains. Therefore, initial MR was carried out in Phaser searching for 2 copies of FN3 domain using residues 461-563 of the PDB-REDO model of 2V5Y. To model the FN4 domain, an alignment using HHPRED identified domains from 2M26, 3UTO and 1WFT that were used to generate an ensemble of models in Sculptor (31). This ensemble was used in Phaser, searching for 2 copies. The resulting solution was refined using autobuster (GlobalPhasing) and manually rebuilt using COOT. This was followed by density modification using Parrot (32) and NCS averaging of autobuster phases followed by autobuilding using Buccaneer (33). Further refinements of all structures were performed using COOT (21), ISOLDE (34) and phenix.refine (35). Graphical figures were rendered in PyMol (Schrödinger LLC). The atomic coordinates and structure factors have been deposited in the Protein Data Bank, www.pdb.org under accession codes 8A1F (PTPRK-MIFN1), 8A16 and 8A17 (PTPRM-FN3-4).

AlphaFold2 Multimer structure predictions

All AF2 models were generated using default parameters and run via a locally installed version of AF2 (36). All models and associated statistics have been deposited in the University of Cambridge Data Repository (https://doi.org/10.17863/CAM.84929).

Small-angle X-ray Scattering

bioRxiv preprint doi: https://doi.org/10.1101/2022.06.23.497309; this version posted June 23, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

SAXS experiments were performed using SEC-SAXS at the EMBL-P12 bioSAXS beam line (PETRAIII, DESY, Hamburg, Germany) (18) with inline MALS, refractive index and UV detectors (Wyatt miniDAWN® TREOS®, a Wyatt Optilab T-rEX (RI) refractometer and Agilent variable wavelength UV-Vis detector recording at 280 nm) (37). The SEC-SAXS-MALS data were recorded as detailed in **Table S2**. For data collection, 40 µL of PTPRK-ECD (2.3 mg/mL) and 30 µL of PTPRM-ECD (7.0 mg/mL) were injected at 0.35 mL/min onto an S200 Increase 5/150 column (Cytiva) equilibrated in 50 mM MES pH 6.0, 250 mM NaCl, 3% (v/v) glycerol. The SAXS data were recorded on a Pilatus 6M detector as a set of 2880 2D-data frames with 0.25 s exposure through the entire column elution. The 2D-to-1D azimuthal averaging was performed using the SASFLOW pipeline (18). The subtraction of appropriate buffer scattering intensities from the sample scattering collected through the single elution peak of either protein was performed using CHROMIXS (18,38). The processing and analysis of the final scaled and averaged SAXS data were performed using the ATSAS 3.0.2 software package (39). The extrapolated forward scattering intensity at zero angle, $I(0)$, and the radius of gyration, R_g, were calculated from the Guinier approximation ($\ln I(s)$ vs s^2, for $sR_g < 1.3$). The radius of gyration of cross section, R_g^c, was estimated using a modified Guinier plot ($\ln(I(s)s)$ vs s^2, in the s^2 range between 0.025–1.0 nm^{-2}), while the dimensionless Kratky plots ($(sR_g)^2 I(s)/I(0)$ vs sR_g) were generated as described in (16). Shape classification was performed using DATCLASS (40). The distance distributions in real-space $p(r)$ were calculated using GNOM (41). A concentration-independent estimate of the molecular weight, extracted directly from the SAXS data, was determined using a Bayesian consensus method (42), while MALS-RI-UV conjugate molecular weight validation was performed as described above (41-43).

Ab initio modeling was performed using multiple individual DAMMIN shape reconstructions followed by spatial alignment and bead-occupancy/volume correction to generate both an averaged spatial representation and a final single dummy atom model for each protein (17,44). Atomistic models for the full ECDs of PTPRM and PTPRK were generated by combining available crystal structures with AF2 models for missing regions. Specifically, for PTPRM the available X-ray structures (PDB ID 2V5Y and the FN3-4 structure determined here) were combined following superposition of the FN3 domains (present in both models). Short stretches of missing residues (19-21

bioRxiv preprint doi: https://doi.org/10.1101/2022.06.23.497309; this version posted June 23, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

and 725-752) were added from the top-ranked AF2 model. For PTPRK, the MIFN1 X-ray structure determined here replaced the equivalent region in the top AF2 model. For both ECD models, the C-terminal His_6 tag and hydrogens were added as well as complex mammalian glycans using the GLYCOSYLATION module of ATSAS in combination with the carbohydrate builder from the Glycam server (https://glycam.org/cb/). Subsequent rigid-body modelling of the domain orientations to the SAXS data was performed using CORAL (19). Five rigid bodies were defined: the MAM plus Ig domains together and the four FN domains independently. The final atomistic model fits to the SAXS profiles were calculated using CRYSOL (35 spherical harmonics, 256 points, with constant adjustment and the inclusion of explicit hydrogens) (45) and the quality evaluated using the reduced χ^2 test and CorMap *P*-value. The SEC-SAXS and SEC-MALS data as well as the SAXS data modeling and analysis are made available in the Small Angle Scattering Biological Data Bank (SASBDB) (46), with the accession codes: SASDPF3 (PTPRK-ECD) and SASDPG3 (PTPRM-ECD) (19,45).

Acknowledgements

We acknowledge Diamond Light Source for time on beamlines I03, I04 and I04-1 under proposal MX15916 and MX21426. Remote access was supported in part by the EU FP7 infrastructure grant BIOSTRUCT-X (Contract No. 283570). The synchrotron SAXS data was collected at beamline P12 operated by EMBL Hamburg at the PETRA III storage ring (DESY, Hamburg, Germany) with support from iNEXT-Discovery, project number 871037, funded by the Horizon 2020 programme of the European Commission. We thank Radu Aricescu for sharing DNA constructs of PTPRM. A Titan V graphics card used for this research was donated to S.C.G. by the NVIDIA Corporation. For the purpose of open access, the author has applied a Creative Commons Attribution (CC BY) licence to any Author Accepted Manuscript version arising from this submission.

Funding Information

This study was supported by a Sir Henry Dale Fellowship jointly funded by the Wellcome Trust and the Royal Society awarded to H.J.S. (109407/Z/15/Z). J.E.D. was supported by a Royal Society University Research Fellowship (UF100371) and a

bioRxiv preprint doi: https://doi.org/10.1101/2022.06.23.497309; this version posted June 23, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

Wellcome Trust Senior Research Fellowship (219447/Z/19/Z). I.M.H. was funded by a CIMR PhD studentship and Babraham Institutional bridging funds. E.R.C. was supported by a Royal Society Enhancement Award (RGF\EA\180151) and M.S. was supported by a Wellcome Trust PhD studentship (203984). A.S.A.M. was supported by a Deutsche Forschungsgemeinschaft grant (SV 9/11-1). H.J.S. is an EMBO Young Investigator and Lister Institute Prize Fellow.

Data availability

The atomic coordinates and structure factors have been deposited in the Protein Data Bank, www.pdb.org under accession codes 8A1F (PTPRK-MIFN1), 8A16 and 8A17 (PTPRM-FN3-4). All AF2 models and associated statistics have been deposited in the University of Cambridge Data Repository (https://doi.org/10.17863/CAM.84929). The SEC-SAXS-MALS data and models are available in the Small Angle Scattering Biological Data Bank (SASBDB) (46), with the accession codes: SASDPF3 (PTPRK-ECD) and SASDPG3 (PTPRM-ECD).

Conflict of Interest

The authors declare they have no conflict of interest.

REFERENCES

1. Garcia, M. A., Nelson, W. J., and Chavez, N. (2018) Cell-Cell Junctions Organize Structural and Signaling Networks. *Cold Spring Harb Perspect Biol* **10**, a029181
2. McLachlan, R. W., and Yap, A. S. (2007) Not so simple: the complexity of phosphotyrosine signaling at cadherin adhesive contacts. *J. Mol. Med.* **85**, 545-554
3. Chiasson-MacKenzie, C., and McClatchey, A. I. (2018) Cell-Cell Contact and Receptor Tyrosine Kinase Signaling. *Cold Spring Harb Perspect Biol* **10**, a029215
4. Young, K. A., Biggins, L., and Sharpe, H. J. (2021) Protein tyrosine phosphatases in cell adhesion. *Biochem. J.* **478**, 1061-1083
5. Fearnley, G. W., Young, K. A., Edgar, J. R., Antrobus, R., Hay, I. M., Liang, W.-C., Martinez-Martin, N., Lin, W., Deane, J. E., and Sharpe, H. J. (2019) The homophilic receptor PTPRK selectively dephosphorylates multiple junctional regulators to promote cell–cell adhesion. *eLife* **8**, e44597
6. Zondag, G. C. M., Reynolds, A. B., and Moolenaar, W. H. (2000) Receptor Protein-tyrosine Phosphatase RPTPμ Binds to and Dephosphorylates the Catenin p120 ctn. *J. Biol. Chem.* **275**, 11264-11269

bioRxiv preprint doi: https://doi.org/10.1101/2022.06.23.497309; this version posted June 23, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

7. Anders, L., Mertins, P., Lammich, S., Murgia, M., Hartmann, D., Saftig, P., Haass, C., and Ullrich, A. (2006) Furin-, ADAM 10-, and gamma-secretase-mediated cleavage of a receptor tyrosine phosphatase and regulation of beta-catenin's transcriptional activity. *Mol Cell Biol* **26**, 3917-3934
8. Tonks, N. K. (2006) Protein tyrosine phosphatases: from genes, to function, to disease. *Nat. Rev. Mol. Cell Biol.* **7**, 833
9. Becka, S., Zhang, P., Craig, S. E. L., Lodowski, D. T., Wang, Z., and Brady-Kalnay, S. M. (2010) Characterization of the adhesive properties of the type IIb subfamily receptor protein tyrosine phosphatases. *Cell Commun. Adhes.* **17**, 34-47
10. Gebbink, M. F., Zondag, G. C., Wubbolts, R. W., Beijersbergen, R. L., van Etten, I., and Moolenaar, W. H. (1993) Cell-cell adhesion mediated by a receptor-like protein tyrosine phosphatase. *J. Biol. Chem.* **268**, 16101-16104
11. Gebbink, M. F., Zondag, G. C., Koningstein, G. M., Feiken, E., Wubbolts, R. W., and Moolenaar, W. H. (1995) Cell surface expression of receptor protein tyrosine phosphatase RPTP mu is regulated by cell-cell contact. *J. Cell Biol.* **131**, 251
12. Brady-Kalnay, S. M., Flint, A. J., and Tonks, N. K. (1993) Homophilic binding of PTP mu, a receptor-type protein tyrosine phosphatase, can mediate cell-cell aggregation. *J. Cell Biol.* **122**, 961-972
13. Aricescu, A. R., Siebold, C., Choudhuri, K., Chang, V. T., Lu, W., Davis, S. J., van der Merwe, P. A., and Jones, E. Y. (2007) Structure of a Tyrosine Phosphatase Adhesive Interaction Reveals a Spacer-Clamp Mechanism. *Science* **317**, 1217
14. Aricescu, A. R., Hon, W. C., Siebold, C., Lu, W., van der Merwe, P. A., and Jones, E. Y. (2006) Molecular analysis of receptor protein tyrosine phosphatase μ-mediated cell adhesion. *EMBO J.* **25**, 701-712
15. Zondag, G. C. M., Koningstein, G. M., Jiang, Y.-P., Sap, J., Moolenaar, W. H., and Gebbink, M. F. B. G. (1995) Homophilic Interactions Mediated by Receptor Tyrosine Phosphatases μ and κ. A critial role for the novel extracellular MAM domain. *J. Biol. Chem.* **270**, 14247-14250
16. Receveur-Brechot, V., and Durand, D. (2012) How random are intrinsically disordered proteins? A small angle scattering perspective. *Curr. Protein Pept. Sci.* **13**, 55-75
17. Svergun, D. I. (1999) Restoring low resolution structure of biological macromolecules from solution scattering using simulated annealing. *Biophys. J.* **76**, 2879-2886
18. Blanchet, C. E., Spilotros, A., Schwemmer, F., Graewert, M. A., Kikhney, A., Jeffries, C. M., Franke, D., Mark, D., Zengerle, R., Cipriani, F., Fiedler, S., Roessle, M., and Svergun, D. I. (2015) Versatile sample environments and automation for biological solution X-ray scattering experiments at the P12 beamline (PETRA III, DESY). *J. Appl. Cryst.* **48**, 431-443
19. Petoukhov, M. V., Franke, D., Shkumatov, A. V., Tria, G., Kikhney, A. G., Gajda, M., Gorba, C., Mertens, H. D., Konarev, P. V., and Svergun, D. I. (2012) New developments in the ATSAS program package for small-angle scattering data analysis. *J. Appl. Cryst.* **45**, 342-350
20. consortium, P.-K. (2022) PDBe-KB: collaboratively defining the biological context of structural data. *Nucleic Acids Res* **50**, D534-D542
21. Emsley, P., Lohkamp, B., Scott, W. G., and Cowtan, K. (2010) Features and development of Coot. *Acta Crystallogr. D* **66**, 486-501

bioRxiv preprint doi: https://doi.org/10.1101/2022.06.23.497309; this version posted June 23, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

22. Nagae, M., and Yamaguchi, Y. (2012) Function and 3D structure of the N-glycans on glycoproteins. *Int J Mol Sci* **13**, 8398-8429
23. Gecht, M., von Bülow, S., Penet, C., Hummer, G., Hanus, C., and Sikora, M. (2021) GlycoSHIELD: a versatile pipeline to assess glycan impact on protein structures. *BioRxiv*, doi.org/10.1101/2021.1108.1104.455134
24. Bager, R., Johansen, J. S., Jensen, J. K., Stensballe, A., Jendroszek, A., Buxbom, L., Sorensen, H. P., and Andreasen, P. A. (2013) Protein conformational change delayed by steric hindrance from an N-linked glycan. *J Mol Biol* **425**, 2867-2877
25. Vendome, J., Felsovalyi, K., Song, H., Yang, Z., Jin, X., Brasch, J., Harrison, O. J., Ahlsen, G., Bahna, F., Kaczynska, A., Katsamba, P. S., Edmond, D., Hubbell, W. L., Shapiro, L., and Honig, B. (2014) Structural and energetic determinants of adhesive binding specificity in type I cadherins. *Proc. Natl Acad. Sci. USA* **111**, E4175-4184
26. Wilkins, M. R., Gasteiger, E., Bairoch, A., Sanchez, J. C., Williams, K. L., Appel, R. D., and Hochstrasser, D. F. (1999) Protein identification and analysis tools in the ExPASy server. *Methods Mol Biol* **112**, 531-552
27. Walter, T. S., Mancini, E. J., Kadlec, J., Graham, S. C., Assenberg, R., Ren, J., Sainsbury, S., Owens, R. J., Stuart, D. I., Grimes, J. M., and Harlos, K. (2008) Semi-automated microseeding of nanolitre crystallization experiments. *Acta Crystallogr. F* **64**, 14-18
28. Winter, G., Waterman, D. G., Parkhurst, J. M., Brewster, A. S., Gildea, R. J., Gerstel, M., Fuentes-Montero, L., Vollmar, M., Michels-Clark, T., Young, I. D., Sauter, N. K., and Evans, G. (2018) DIALS: implementation and evaluation of a new integration package. *Acta Crystallogr. D* **74**, 85-97
29. Evans, P. R., and Murshudov, G. N. (2013) How good are my data and what is the resolution? *Acta Crystallogr. D* **69**, 1204-1214
30. McCoy, A. J. (2007) Solving structures of protein complexes by molecular replacement with Phaser. *Acta Crystallogr. D* **63**, 32-41
31. Bunkoczi, G., and Read, R. J. (2011) Improvement of molecular-replacement models with Sculptor. *Acta Crystallogr. D* **67**, 303-312
32. Griffith, D., and Holehouse, A. S. (2021) PARROT is a flexible recurrent neural network framework for analysis of large protein datasets. *eLife* **10**
33. Cowtan, K. (2006) The Buccaneer software for automated model building. 1. Tracing protein chains. *Acta Crystallogr. D* **62**, 1002-1011
34. Croll, T. I. (2018) ISOLDE: a physically realistic environment for model building into low-resolution electron-density maps. *Acta Crystallogr. D* **74**, 519-530
35. Afonine, P. V., Grosse-Kunstleve, R. W., Echols, N., Headd, J. J., Moriarty, N. W., Mustyakimov, M., Terwilliger, T. C., Urzhumtsev, A., Zwart, P. H., and Adams, P. D. (2012) Towards automated crystallographic structure refinement with phenix.refine. *Acta Crystallogr. D* **68**, 352-367
36. Jumper, J., Evans, R., Pritzel, A., Green, T., Figurnov, M., Ronneberger, O., Tunyasuvunakool, K., Bates, R., Zidek, A., Potapenko, A., Bridgland, A., Meyer, C., Kohl, S. A. A., Ballard, A. J., Cowie, A., Romera-Paredes, B., Nikolov, S., Jain, R., Adler, J., Back, T., Petersen, S., Reiman, D., Clancy, E., Zielinski, M., Steinegger, M., Pacholska, M., Berghammer, T., Bodenstein, S., Silver, D., Vinyals, O., Senior, A. W., Kavukcuoglu, K., Kohli, P., and Hassabis, D. (2021) Highly accurate protein structure prediction with AlphaFold. *Nature* **596**, 583-589
37. Gräwert, M. A., Da Vela, S., Gräwert, T. W., and Molodenskiy, D. S. (2020) Adding Size Exclusion Chromatography (SEC) and Light Scattering (LS) Devices

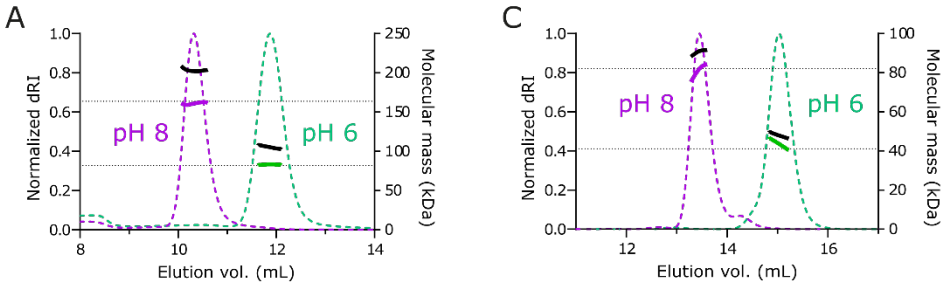
bioRxiv preprint doi: https://doi.org/10.1101/2022.06.23.497309; this version posted June 23, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

to Obtain High-Quality Small Angle X-Ray Scattering (SAXS) Data. *Crystals* **10**, 975

38. Panjkovich, A., and Svergun, D. I. (2018) CHROMIXS: automatic and interactive analysis of chromatography-coupled small-angle X-ray scattering data. *Bioinformatics* **34**, 1944-1946
39. Manalastas-Cantos, K., Konarev, P. V., Hajizadeh, N. R., Kikhney, A. G., Petoukhov, M. V., Molodenskiy, D. S., Panjkovich, A., Mertens, H. D. T., Gruzinov, A., Borges, C., Jeffries, C. M., Svergun, D. I., and Franke, D. (2021) ATSAS 3.0: expanded functionality and new tools for small-angle scattering data analysis. *J. Appl. Cryst.* **54**, 343-355
40. Franke, D., Jeffries, C. M., and Svergun, D. I. (2018) Machine Learning Methods for X-Ray Scattering Data Analysis from Biomacromolecular Solutions. *Biophys. J.* **114**, 2485-2492
41. Svergun, D. I. (1992) Determination of the regularization parameter in indirect-transform methods using perceptual criteria. *J. Appl. Cryst.* **25**, 495-503
42. Hajizadeh, N. R., Franke, D., Jeffries, C. M., and Svergun, D. I. (2018) Consensus Bayesian assessment of protein molecular mass from solution X-ray scattering data. *Sci Rep* **8**, 7204
43. Semenyuk, A. V., and Svergun, D. I. (1991) GNOM– a program package for small-angle scattering data processing. *J. Appl. Cryst.* **24**, 537-540
44. Volkov, V. V., and Svergun, D. I. (2003) Uniqueness of ab-initio shape determination in small-angle scattering. *J. Appl. Cryst.* **36**, 860-864
45. Svergun, D. I. B., C.; Koch, M. H. J. . (1995) CRYSOL – a Program to Evaluate X-ray Solution Scattering of Biological Macromolecules from Atomic Coordinates. . *J. Appl. Cryst.* **28**, 768-773
46. Kikhney, A. G., Borges, C. R., Molodenskiy, D. S., Jeffries, C. M., and Svergun, D. I. (2020) SASBDB: Towards an automatically curated and validated repository for biological scattering data. *Protein Sci* **29**, 66-75

bioRxiv preprint doi: https://doi.org/10.1101/2022.06.23.497309; this version posted June 23, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

Figures and Legends



B

Construct	Total molecular mass (kDa)	Molecular mass protein (kDa)	Predicted mass (kDa)
FL pH 6	105.4 ± 0.5	83.2 ± 0.4	81.9 (monomer)
FL pH 8	203.1 ± 1.0	161.1 ± 0.8	163.8 (dimer)
MIFN1 pH 6	47.9 ± 0.1	43.3 ± 0.2	41.1 (monomer)
MIFN1 pH 8	90.6 ± 0.3	81.3 ± 0.2	82.2 (dimer)

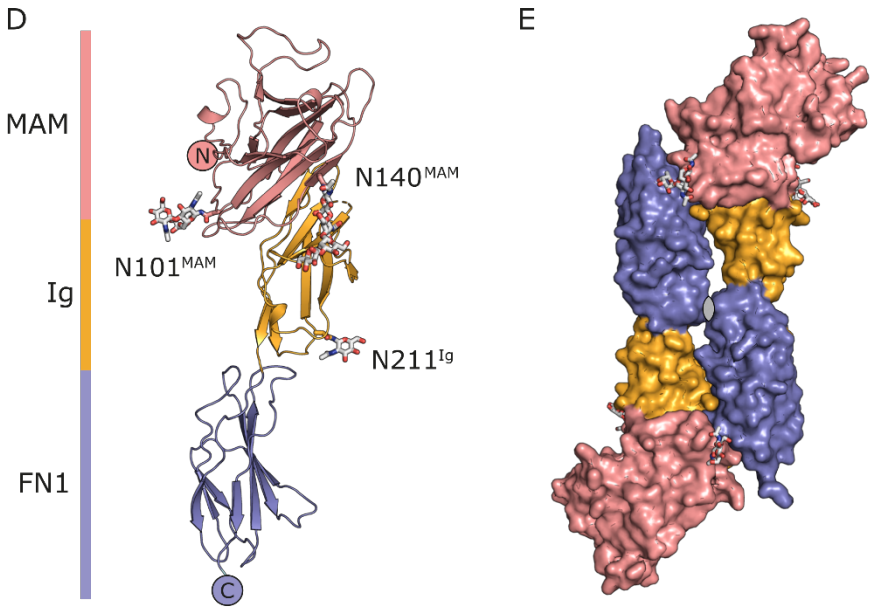


Figure 1. Structure of the PTPRK extracellular domain minimal dimerization unit.

(A) SEC-MALS analysis of the full-length PTPRK-ECD. The SEC elution profile (normalized differential refractive index, dRI, dashed colored lines) and molecular mass distribution (total mass; black solid line, protein mass; colored solid line) at pH 8 (magenta) and pH 6 (green) are shown. Dashed horizontal lines indicated the predicted protein mass of PTPRK-ECD monomer and dimer. **(B)** Summary table of SEC-MALS mass determinations with and without conjugate analysis for attached glycans. **(C)** SEC-MALS analysis of the PTPRK-MIFN1 fragment, illustrated as described for A. **(D)** Ribbon diagram of the PTPRK-MIFN1 monomer, colored by domain. MAM: pink, Ig: orange, FN1: blue. N-linked glycosylation sites are labelled

bioRxiv preprint doi: https://doi.org/10.1101/2022.06.23.497309; this version posted June 23, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

and glycans shown in stick representation. **(E)** Surface representation of the PTPRK-MIFN1 dimer, colored as in C. The 2-fold non-crystallographic symmetry axis is marked.

bioRxiv preprint doi: https://doi.org/10.1101/2022.06.23.497309; this version posted June 23, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

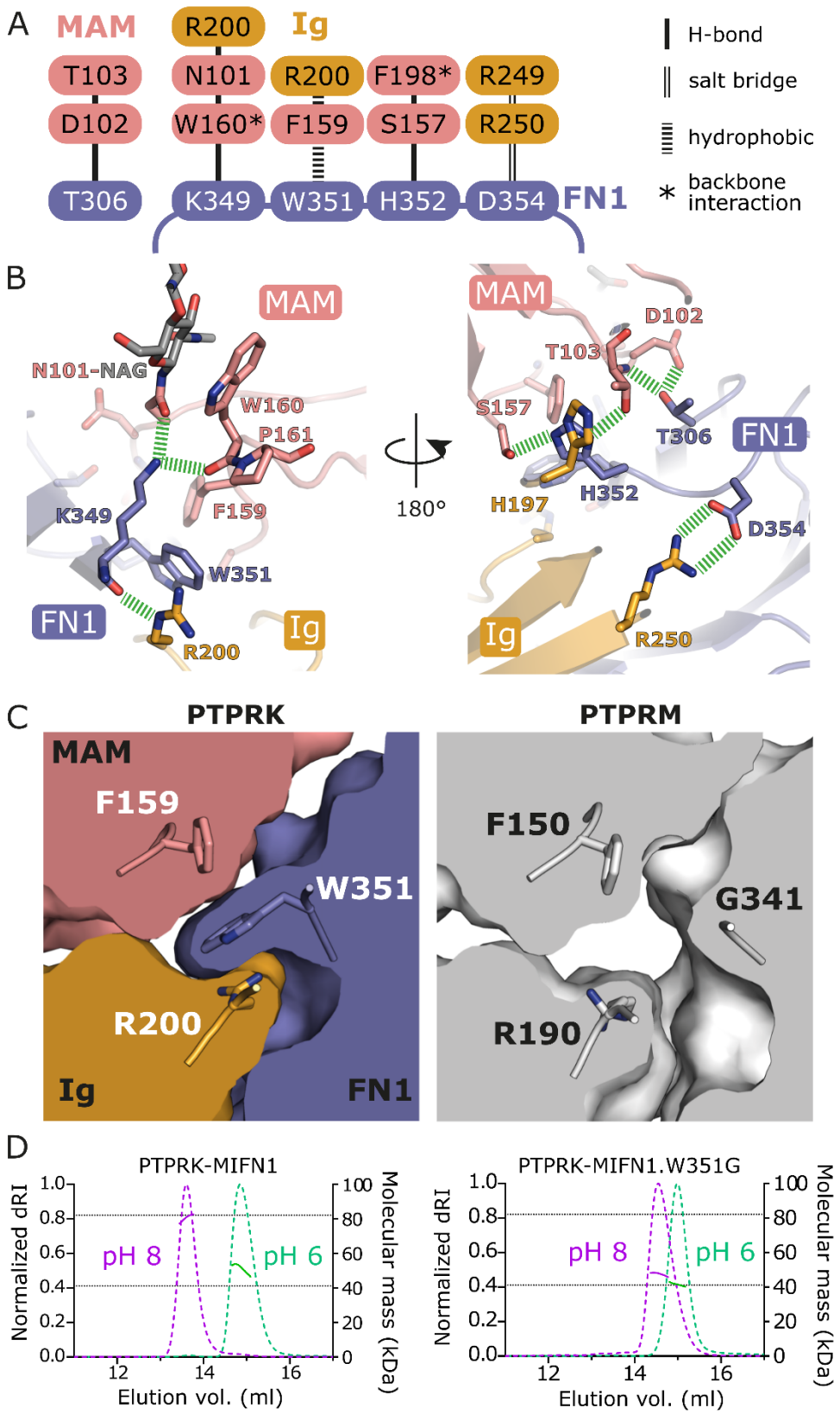


Figure 2. Dimerization interface of the PTPRK-MIFN1 structure

bioRxiv preprint doi: https://doi.org/10.1101/2022.06.23.497309; this version posted June 23, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

(A) Schematic diagram showing the hydrogen bonds, salt bridges and hydrophobic interactions at the MAM-Ig:FN1 interface within the PTPRK-MIFN1 homodimer. Stacked residues on the MAM-Ig side denote residues which have a common interacting residue in the FN1 domain. *denotes interactions formed by peptide backbone carbonyl oxygens. **(B)** Molecular detail of the PTPRK-MIFN1 homodimer interface. Side chains and backbone atoms of key interaction residues from (A) are highlighted, with hydrogen bonds and salt bridges highlighted (dashed green lines). The N-acetylglucosamine of the N101 linked glycan chain is shown (gray sticks). **(C)** Left: Cross-section of the MAM-Ig:FN1 interface in the PTPRK-MIFN1 dimer. W351 in the FN1 domain forms hydrophobic interactions with F159 and the aliphatic portion of R200 in the MAM and Ig domains, respectively. Right: Cross-section of the MAM-Ig:FN1 interface in the PTPRM-ECD dimer (PDB ID: 2V5Y) shown in the same orientation. **(D)** Left: SEC-MALS analysis of WT PTPRK-MIFN1. The SEC elution profile (normalized differential refractive index, dRI, dashed colored lines) and calculated protein mass distributions following conjugate analysis (colored solid line) are shown at pH 8 (magenta) and pH 6 (green). Dashed horizontal lines indicated the predicted protein mass of PTPRK-MIFN1 monomer and dimer. Right: SEC-MALS analysis of PTPRK-MIFN1.W351G, presented as for WT PTPRK-MIFN1 (left panel).

bioRxiv preprint doi: https://doi.org/10.1101/2022.06.23.497309; this version posted June 23, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

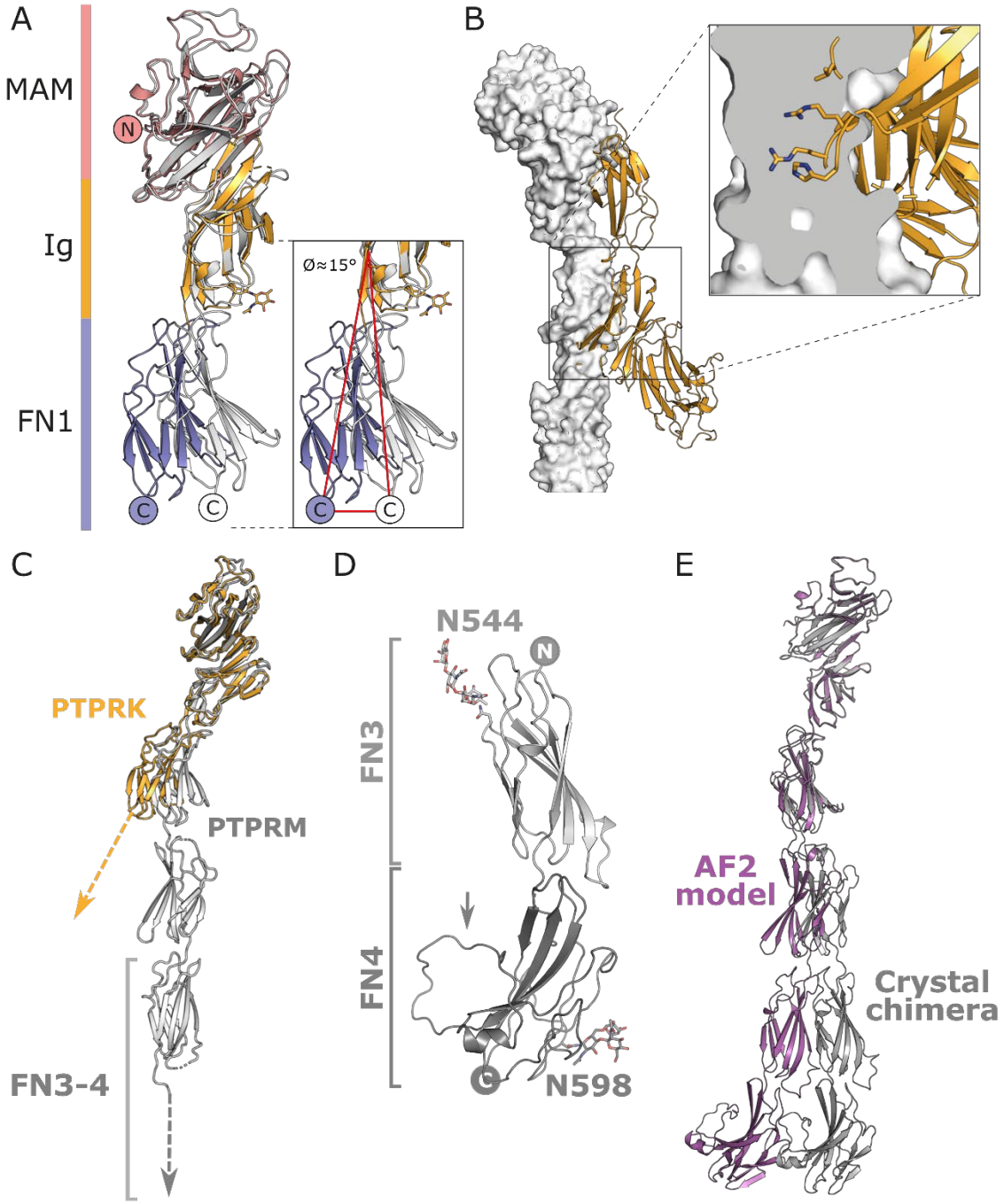


Figure 3. The PTPRK and PTPRM ECD structures possess different inter-domain conformations.

(A) Ribbon diagram of a PTPRK-MIFN1 monomer, colored by domain. MAM: pink, Ig: orange, FN1: blue. Shown is a structural alignment over the MAM-Ig domains of the equivalent MIFN1 fragment from the full-length PTPRM-ECD structure (gray, PDB ID: 2V5Y). Inset: The FN1 domain of PTPRK possesses a ~15° shift from a hinge point at the Ig-FN1 boundary when compared to PTPRM. The glycan on PTPRK N211, which is absent in PTPRM, is shown (orange sticks). **(B)** The full-length PTPRM-ECD (gray

bioRxiv preprint doi: https://doi.org/10.1101/2022.06.23.497309; this version posted June 23, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

surface) was aligned to one chain of the PTPRK-MIFN1 dimer (orange cartoon). Inset: Cross-section of the PTPRM-ECD overlay highlighting the extensive steric clashes when aligning PTPRM to the PTPRK dimer. **(C)** Ribbon diagram of a PTPRK-MIFN1 monomer (orange) aligned to the MAM-Ig domains of the PTPRM-ECD (gray). Dashed arrows highlight the potential for long-range structural divergence as a result of differing hinge angles at the Ig-FN1 boundary. **(D)** Ribbon diagram of the PTPRM-FN3-4 structure. N-linked glycosylation sites are labelled and glycans shown in stick representation. The loop in FN4 containing the furin cleavage site is labelled (arrow). (E) Comparison of structures for the full-length ECD of PTPRM determined using available crystal structures (chimera, gray) and AlphaFold2 (AF2, purple). Structures were overlaid using the MAM-Ig domains only.

bioRxiv preprint doi: https://doi.org/10.1101/2022.06.23.497309; this version posted June 23, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

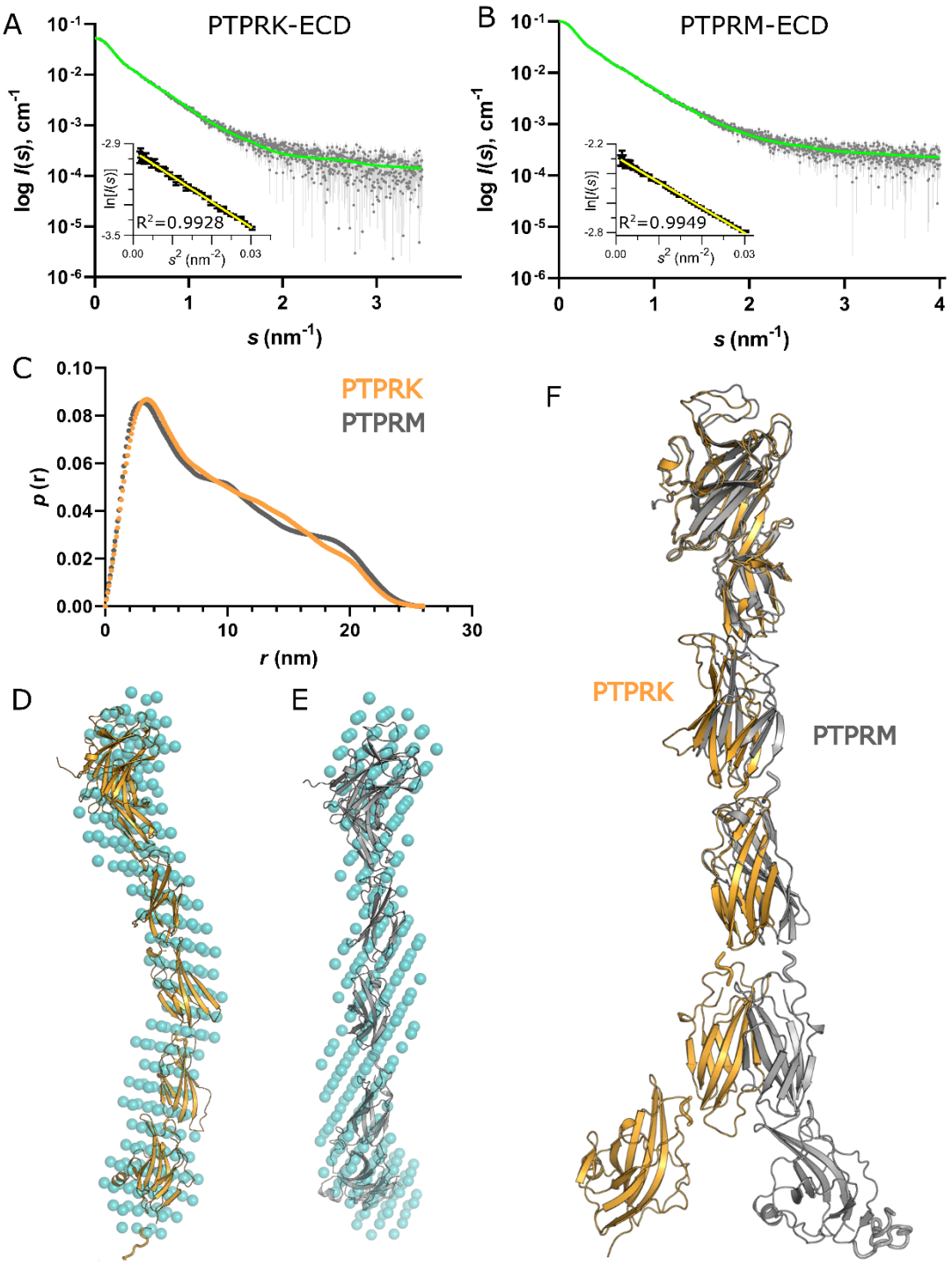


Figure 4. Small angle X-ray scattering data and ECD modelling.

Averaged SEC-SAXS profile of **(A)** PTPRK-ECD and **(B)** PTPRM-ECD (gray diamonds) and corresponding fit against the data of respective pseudo-atomic models shown in (**D**) and (**E**) (green line). *Inset*, Guinier plot of ln*I*(*s*) versus s^2 for $sR_g < 1.3$. **(C)** *p*(*r*) versus *r* profiles of PTPRK-ECD (orange) and PTPRM-ECD (gray)**. (D)** The *ab initio* dummy-atom bead model of PTPRK-ECD calculated using DAMMIN (cyan spheres) overlaid with the best pseudo-atomic model (lowest χ^2) generated by fitting

bioRxiv preprint doi: https://doi.org/10.1101/2022.06.23.497309; this version posted June 23, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

to the SAXS data. **(E)** As for panel D but for PTPRM-ECD data. **(F)** A spatial comparison of the PTPRK-ECD (orange) and PTPRM-ECD (gray) SAXS-based models aligned relative to the N-terminal MAM-Ig domains.