BIOCHEMISTRY
including biophysical chemistry & molecular biology
Article

Biochemistry **2010,** *49,* 7131–7150 7131
DOI: 10.1021/bi100687f

Human Plasminogen Kringle 3: Solution Structure, Functional Insights, Phylogenetic Landscape[†,‡]

Martin T. Christen,[§] Pascal Frank,[∥] Johann Schaller,[∥] and Miguel Llinás*[,§]

[§]*Department of Chemistry, Carnegie Mellon University, Pittsburgh, Pennsylvania 15213, and* [∥]*Department of Chemistry and Biochemistry, University of Bern, CH-3012 Bern, Switzerland*

Received May 6, 2010; Revised Manuscript Received June 18, 2010

ABSTRACT: Human plasminogen kringle 3 (hPgn K3) domain contains most elements of the canonical lysine-binding site (LBS) found in other Pgn kringles. However, it does not exhibit affinity for either lysine or structurally related zwitterionic ligands. It has been shown that lysine-binding activity can be engineered via a Lys57 → Asp mutation [Bürgin, J., and Schaller, J. (2009) *Cell. Mol. Life Sci. 55*, 135]. Using a recombinant construct expressed in *Escherichia coli*, the three-dimensional solution structure of hPgn K3 was determined via NMR spectroscopy [heavy atom averaged rmsd = 0.35 ± 0.07 Å (backbone) and 0.75 ± 0.12 Å (all)]. The $^{1}H/^{15}N$ heteronuclear single-quantum correlated (HSQC) spectra for both wild-type K3 and mutated [r(K57D)K3] structures are essentially identical, implying that the two structures are effectively isomorphous. The affinity of r(K57D)K3 for the lysine analogue *trans*-(aminomethyl)cyclohexanecarboxylic acid (AMCHA) was investigated from ligand-induced NMR chemical shift perturbations, which enabled for mapping the binding site on the mutated domain surface. The equilibrium association constant, K_a, was determined to be ∼5.23 ± 0.03 mM^{-1}. Homology modeling combined with *in silico* docking of lysine-like zwitterionic ligands via AutoDock 4.0 supports functionality of the engineered (K57D)K3 LBS, whose electrostatic focal centers are defined by the Arg36/Arg71 cationic and Asp55/Asp57 anionic pairs. Comparison of K3-type sequences from different vertebrates, including kringles from hedgehog apolipoprotein(a) [Apo(a)] and Apo(a)-related (Arp) sequences, reveals that Lys57 is confined to the hPgn variant. Based on the likely phylogeny and ligand affinities of the homologous domains, it is suggested that the hPgn K3 is unique in that all other K3-type domains, including hedgehog Apo(a) and all Arp domains, except K3(1), are predicted to variously exhibit lysine-binding capability. In Arp K3(1) an Arg residue fills site 72, replacing the key aromatic residue found in other kringles, thus interfering with a requisite kringle−ligand hydrophobic interaction.

Plasminogen (Pgn),[1] the inactive precursor of the extracellular proteinase plasmin (Plm), is a main component of the blood fibrinolytic cascade (*1*). Besides its role in vascular hemostasis, Pgn is crucial in processes leading to tissue remodeling, such as cell migration and embryogenesis, as well as in tumor growth and metastasis (*2*, *3*). Human Pgn (hPgn) is structured as a mosaic glycoprotein of 791 amino acid residues encompassing 7 globular folding units (*4*): a relatively flexible N-terminal PAN domain of 77 amino acids (*5*, *6*) that is cleaved off during the activation process, an in-tandem array of five homologous kringle[2] (K) repeats of ∼80 amino acids each, and a C-terminal serine protease zymogen domain of 231 amino acids.

Sequence alignment of hPgn kringles reveals 35% sequence identity among the 5 repeats, reaching 50% sequence identity within the K1, K2, K3, and K4 set (*7*) (Scheme 1A). Both solution NMR and X-ray crystallographic structures of hPgn K1 (*8*−*10*), K2 (*11*, *12*), K4 (*13*, *14*), and K5 (*15*, *16*) are known. Moreover, the structures of the kringles in tissue-type (*17*−*19*) and urokinase (*20*, *21*) Pgn activators (tPA and uPA, respectively) have been solved; additionally, the X-ray structure of a recombinant form of the in-tandem multidomain construct K1−K2−K3 (angiostatin) has also been determined at 1.75 Å resolution (*22*).

[†]This research was supported by National Institutes of Health Grant HL29409 (M.Ll.) and SNF Grant 3100-039298 (J.S.).

[‡]Coordinates for rK3 have been deposited in the Protein Data Bank: accession code 210S. ^{1}H and ^{15}N chemical shift assignments and experimental NMR restraints for rK3 have been deposited in the Biological Magnetic Resonance Bank: accession code 17062. ^{1}H and ^{15}N chemical shift assignments for r(K57D)K3 have been deposited in the Biological Magnetic Resonance Bank: accession code 17063.

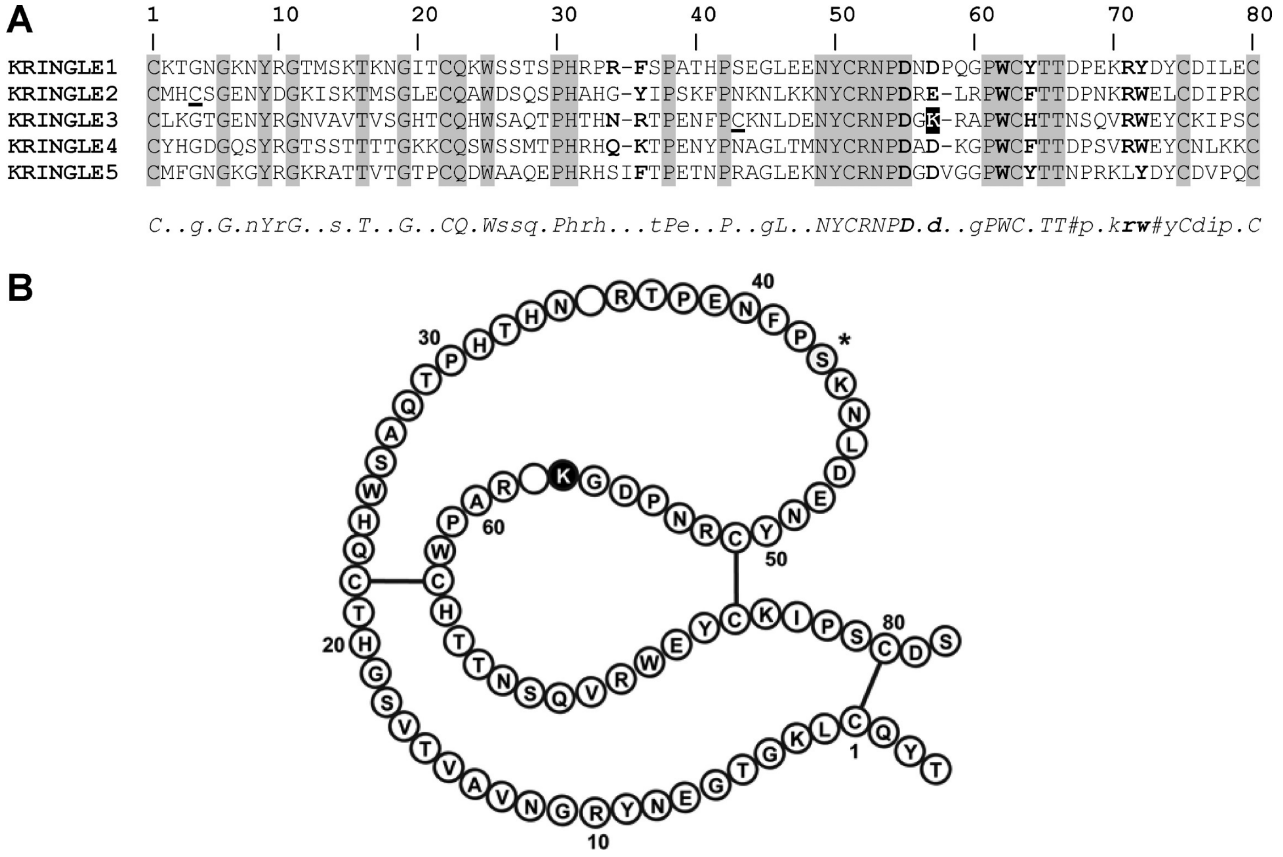
*To whom correspondence should be addressed. E-mail: llinas@andrew.cmu.edu. Telephone: (412) 268-3140. Fax: (412) 268-1061.

[1]Abbreviations: 6-AHA, 6-aminohexanoic acid; AcLys, N^{α}-acetyl-L-lysine; Apo, apolipoprotein; AMCHA, *trans*-(aminomethyl)cyclohexanecarboxylic acid; Arp, hedgehog apolipoprotein(a)-related protein-1; BASA, *p*-benzylaminesulfonic acid; COSY, correlation spectroscopy; DANRE, *Danio rerio* (Zebrafish); Δδ, chemical shift perturbations; Δ*G**, free energy of binding (kcal/mol) estimated via AutoDock 4.0; Δ*p*, weighted average fraction of ligand-bound kringle; ERIEU, *Erinaceus europaeus* (Western European hedgehog); HNHA, ^{15}N-edited $^{1}H^{N}/^{1}H^{\alpha}$ homonuclear *J*-correlated experiment; HNHB, heteronuclear *J*-correlated $^{15}N-^{1}H^{\beta}$ experiment; hPgn, human Pgn; HSQC, heteronuclear single-quantum correlated spectroscopy; K, kringle domain; K_a, kringle−ligand equilibrium association constant; K_a^*, K_a calculated from Δ*G**; $\overline{K_a^*}$, ligand-averaged K_a^*; *l*, dipole length; LBS, lysine-binding site; MACEU, *Macropus eugenii* (Tammar wallaby); MACMU, *Macaca mulatta* (Rhesus macaque); NOE, nuclear Overhauser effect; NOESY, NOE correlation spectroscopy; NOESY-HSQC, 3D ^{15}N-edited NOESY; ONCMY, *Oncorhynchus mykiss* (rainbow trout); ORYLA, *Oryzias latipes* (Japanese ricefish); PAN, preactivation N-terminal domain of hPgn (Glu1−Lys77); PETMA, *Petromyzon marinus* (sea lamprey); Pgn, plasminogen; pH*, glass electrode pH reading, uncorrected for ^{2}H isotope effects; Plm, plasmin; PONAB, *Pongo abelii* (Sumatran orangutan); rK3, recombinant kringle construct TYQ[$K3_{hPgn}$/C297S]DS; r(K57D)K3, recombinant kringle construct TYQ[$K3_{hPgn}$/C297S/K311D]DS; rmsd, root-mean-square deviation; [S], free ligand concentration; TOCSY, total correlated spectroscopy; TOCSY-HSQC, 3D ^{15}N-edited TOCSY; τ_{mix}, NOESY mixing time; TSP, 2,2,3,3-tetradeuterio-3-(trimethylsilyl)-propionic acid; UPGMA, unweighted pair-group method with arithmetic mean.

[2]Unless otherwise specified, henceforth K always refers to hPgn kringles.

Published on Web 07/09/2010

Scheme 1: Primary Structures of Human Plasminogen Kringle Domains[a]



[a]Residue numbering is based on the hPgn K5 sequence (*50*). Lys57 in K3 is denoted in white over black. (A) Sequence alignment of the five domains. Strictly conserved residues are shaded gray; residues located at the canonical LBS are in bold type; free cysteines are underlined. The consensus sequence (35% identity) is in italics; homology ≥90% is in upper case; homology ≥50% (but <90%) is in lower case. Sites occupied by polar residues are denoted by #. (B) Outline of the investigated recombinant hPgn K3. Except for the N- and C-terminal tails (tri- and di-peptide, respectively) and the C43S mutation (*) introduced to avoid dimerization, the construct corresponds to the wild type. Cystine bridges are indicated. Deletions relative to hPgn K5 are shown blank (empty circles).

Each kringle unit contains 3 strictly conserved cystines that generate a characteristic 1−6, 2−4, 3−5 bridged pattern (*23*, *24*). Conformationally, these internal links constrain the kringle to adopt a characteristic fold of dimension ~30 Å × 37 Å × 25 Å (*8*−*16*, *22*), consisting of three loops surrounding a central Cys22−Cys63/Cys51−Cys75 "in cross" disulfide array[3] which, in planar projection, leads the domains to outline a pretzel-like contour (Scheme 1B). An additional, interkringle, cystine bridge covalently links K2 to K3, thus clamping the two repeats to generate a unique two-kringle "supermodule" (*4*).

Kringle domains mediate intermolecular protein−protein interactions (*25*, *26*). In Plm, kringles contribute substrate specificity to the proteinase (*27*). This is achieved via a lysine-binding site (LBS) that enables the domains to preferentially interact with C-terminal Lys residues which, in turn, are generated by the activity of the trypsin-type protease domain. Furthermore, a variety of Lys-like zwitterionic compounds are known to bind to kringles, with some of these affording potent antifibrinolytic drugs (*28*) as they interfere with Pgn binding to the fibrin meshwork of blood clots. In order of decreasing affinities, the specificity of hPgn kringles toward ω-aminocarboxylic ligands ranks K1≳K4≳K5≳K2 (*29*), with K3 showing no detectable binding (*30*, *31*). Hence, K3 is rather unique in that its *modus operandi* remains largely unknown.

The basic scaffolding of the kringles' LBS exposes a hydrophobic groove, axially delimited by electrostatic opposite polar groups contributed by acidic residues at sites 55 and 57 and basic residues at sites 34 (in K1, K4) and 71. Side chains of these residues provide the anionic and cationic centers that complement the electrostatic dipolar configuration of the zwitterionic ligands, thus stabilizing the binding (*23*). Interestingly, position 57 in hPgn K3 is occupied by a Lys residue (Scheme 1), which would be expected to adversely affect the binding of Lys-like zwitterions. Indeed, a measurable affinity of the hPgn K3 for such ligands results from an engineered K57D mutation (*32*). While this affords strong evidence for a potentially functional binding site in hPgn K3, the interaction of the mutant kringle [henceforth (K57D)K3] with zwitterionic ligands has not been mapped onto the protein surface.

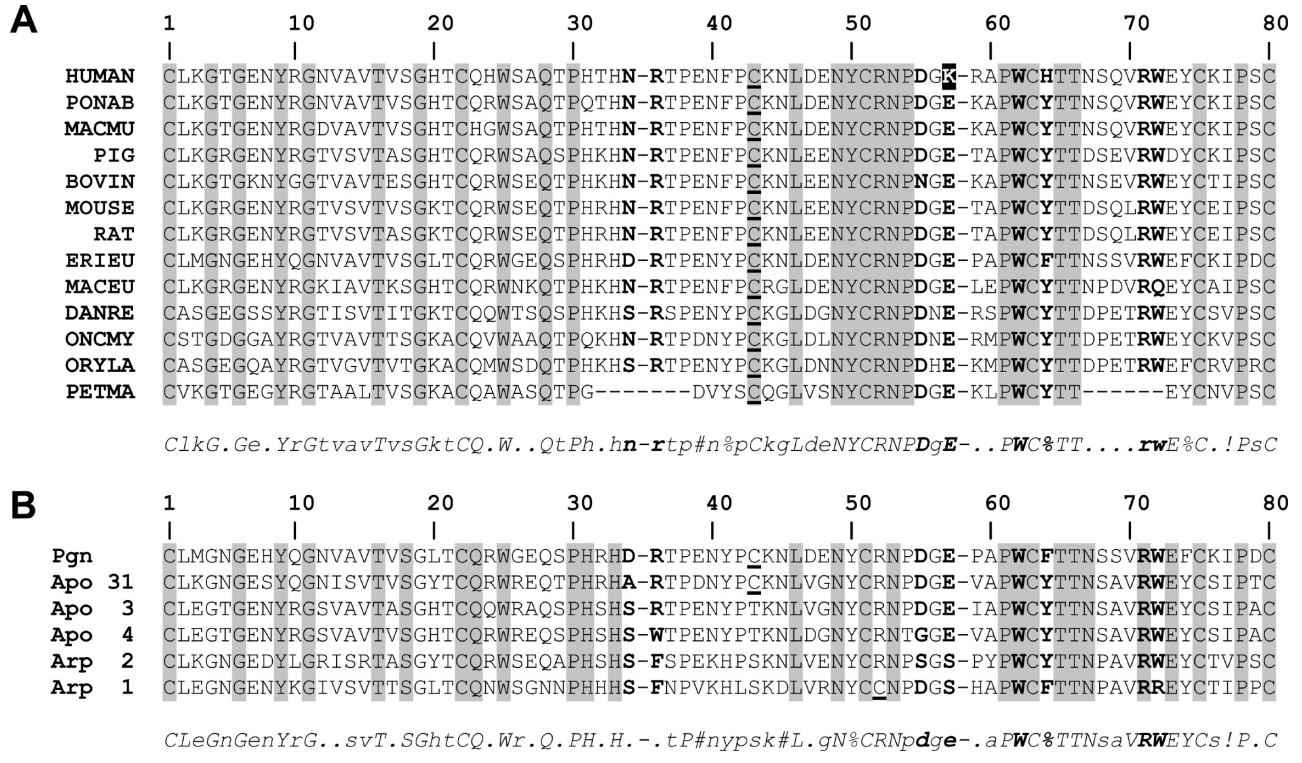
In this paper, we describe the NMR solution structure of a recombinant hPgn K3 construct (Scheme 1B). We also report on the specifics of the ligand-binding properties of (K57D)K3 as well as on the predicted structural/functional implications of K3 variants present in several phylogenetically related sequences (Scheme 2).

EXPERIMENTAL PROCEDURES

Sample Preparation. A recombinant expression vector for hPGn K3 was generated based on the cDNA sequence of

[3]We follow the standard kringle amino acid numbering convention (*50*). Throughout the text, individual residues and naturally occurring substitutions are referred to by three-letter code while one-letter code is used for engineered mutations.

Scheme 2: Sequence Alignment of Kringle 3 Homologues[a]



[a]Residue numbering is based on the hPgn K5 sequence (*50*). Strictly conserved residues are shaded gray; residues located at the canonical LBS are in bold type; free cysteines are underlined. For each group, sequences are ranked according to homology using ClustalW score (*89*) by reference to the top sequence. (A) Pgn K3 orthologues from various species. Labels indicate the organism (*103*) and Uniprot access number (*104*): HUMAN, *Homo sapiens* (P00747); PONAB, *Pongo abelii* (Sumatran orangutan, Q5R8X6); MACMU, *Macaca mulatta* (Rhesus macaque, P12545); PIG, *Sus scrofa* (P06867); BOVIN, *Bos taurus* (P06868); MOUSE, *Mus musculus* (P20918); RAT, *Rattus norvegicus* (Q01177); ERIEU, *Erinaceus europaeus* (Western European hedgehog, Q29485); MACEU, *Macropus eugenii* (Tammar wallaby, O18783); DANRE, *Danio rerio* (Zebrafish, Q6PBA6); ONCMY, *Oncorhynchus mykiss* (Rainbow trout, Q5DVP8); ORYLA, *Oryzias latipes* (Japanese ricefish, Q50LG6); PETMA, *Petromyzon marinus* (Sea lamprey, P33574). Lys57 in the HUMAN sequence is denoted in white over black. (B) ERIEU K3 domain variants from Pgn (5 kringles), Apo(a) (Apo, 31 kringles), and Arp (7 kringles). The sequential kringle number is shown. The third Apo(a) K3 [Apo K3(3)] is identical to *odd*-numbered kringles 5, 7, 9, 11, 13, 15, 17, and 19; Apo K3(4) is identical to *even*-numbered kringles 6, 10, 12, 16, 18, and 20. Arp kringles K3(2) and K3(1) are the most and least homologous to ERIEU Pgn K3, respectively. Highlighted residues are strictly conserved over all 43 represented kringle sequences. Consensus sequences for (A) and (B) (34.62% identity in both cases) are in italics; homology ≥90% is in upper case; homology ≥50% (but <90%) is in lower case. Sites occupied by polar, aromatic, and aliphatic residues are denoted by #, %, and !, respectively.

hPgn (*33*). The construct rK3, which encompasses the region rTYQ[$K3_{hPgn}$(C297S)]DS, contains a C297S mutation to avoid interkringle disulfide bridge formation. An expression vector was similarly generated for the mutated (K57D)K3 hPgn kringle, resulting in the construct rTYQ[$K3_{hPgn}$(C297S/K311D)]DS. Both constructs were expressed in *Escherichia coli* strain M15 (F^- Str^R $lacZ^{del}$) (*34*) and isolated, refolded, and purified following previously published protocols (*30*, *32*); ^{15}N-labeled protein was obtained by growing the *E. coli* cells in M9 minimal medium containing [^{15}N]ammonium chloride as described (*35*). The recombinant proteins were characterized via electrospray ionization mass spectrometry, amino acid analysis, and N-terminal amino acid sequencing by Edman degradation.

For the NMR experiments, uniformly ^{15}N-labeled rK3 and r(K57D)K3 were dissolved in 500 μL of H_2O/D_2O (9:1) at pH* 5.5 to final concentrations of ~1.0 mM. For studies in ^2H_2O, the samples were exchanged thrice with 99.9% ^2H_2O and then dissolved in 500 μL of 99.96% ^2H_2O (Isotec, Inc.), and the pH* was adjusted to 5.7.

NMR Spectroscopy. NMR spectra were recorded at 25 °C on Bruker Advance DRX spectrometers. Homonuclear two-dimensional (2D) COSY (*36*), TOCSY (*37*), and NOESY (*38*) experiments were acquired at 500 MHz using an isotopic natural abundance sample, while ^{15}N-edited 2D HSQC (*39*) and 3D TOCSY-HSQC, NOESY-HSQC (*40*), HNHA (*41*), and HNHB (*42*) data were obtained at 600 MHz with the ^{15}N-labeled material. Water suppression was achieved via presaturation (COSY) or WATERGATE (*43*). TOCSY experiments using MLEV-17 (2D) (*44*) or DIPSI-2 (3D) (*45*) sequences were recorded with mixing times (τ_{mix}) of 20 and 80 ms. 2D NOESY experiments were recorded at τ_{mix} = 100, 200, and 300 ms and 3D HSQC-NOESY with τ_{mix} = 50 and 100 ms.

Data were processed using the FELIX 98 software package (Molecular Simulations, Inc.). The 2D homonuclear and HSQC matrices were 1024 × 1024 real points for the f_1 and f_2 dimensions, respectively. The matrix size for 3D NOESY-HSQC and TOCSY-HSQC was 512 × 128 × 512 for f_1, f_2, and f_3 dimensions while the HNHA and HNHB matrices were 128 × 128 × 256 real points. ^{1}H chemical shifts are by reference to internal *p*-dioxane 3.75 ppm downfield from 2,2,3,3-tetradeuterio-3-(trimethylsilyl)-propionic acid (TSP) (*46*). The nitrogen dimensions are referenced using the default values provided by the Bruker software.

Resonance Assignments and Angular Constraints. Sequential ^{1}H and ^{15}N NMR assignments of rK3 were determined at pH* 5.7 following standard protocols (*47*) that include ^{15}N-edited double resonance experiments (*47*−*49*). The ^{1}H/^{15}N HSQC of rK3 (Figure 1) exhibits good dispersion and allows

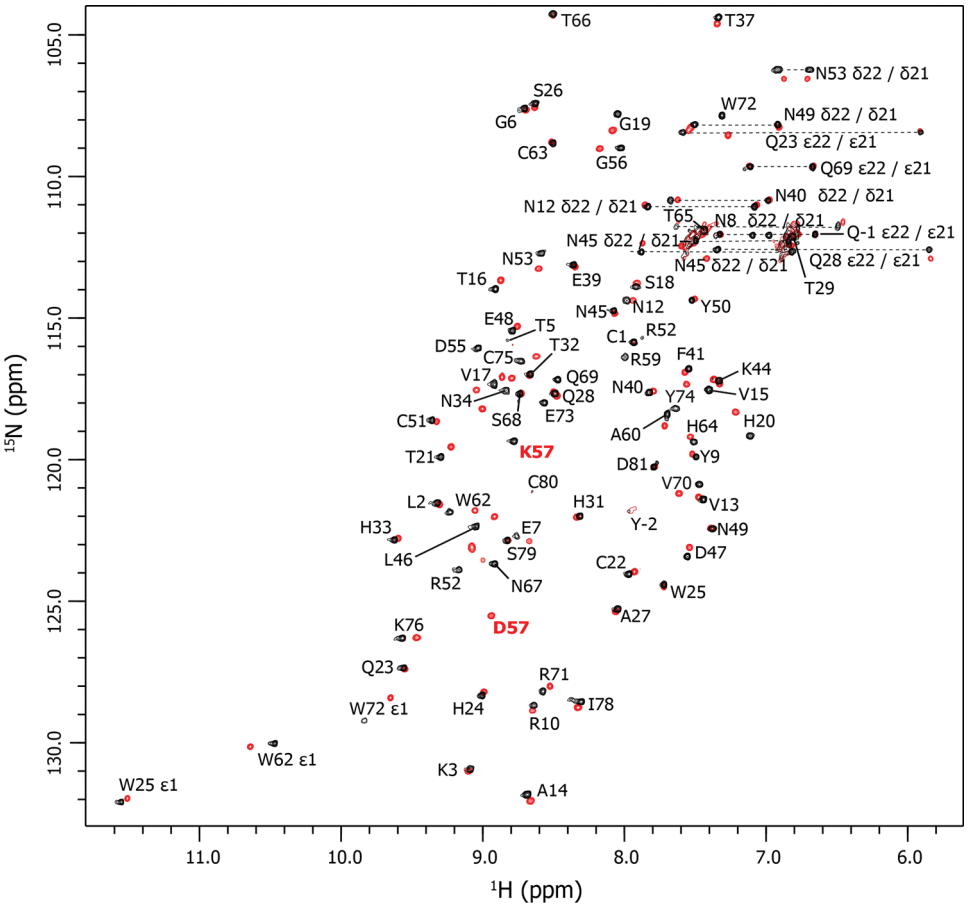


FIGURE 1: Superposed NMR ^{1}H/^{15}N HSQC spectra of the wild-type recombinant human plasminogen kringle 3 and the K57D mutant, cross-peaks in black and red, respectively. Signals arising from Asn $H^{\delta 2}$/$H^{\delta 22}$ and Gln $H^{\varepsilon 21}$/$H^{\varepsilon 22}$ geminal proton pairs are connected by dashed lines. The backbone amide ^{15}N resonance of the mutated residue (red label) appears at δ = 124.9 ppm (Asp57) compared to the corresponding wild-type signal at δ = 119.4 ppm (Lys57). Data were recorded at 600 MHz (^{1}H) and pH* 7.0.

for backbone ^{1}H and ^{15}N resonances to be assigned; the Gly11 ^{15}N backbone resonance is missing because of the H^{N} upfield shift at 4.645 ppm. Side chain ^{1}H signals were assigned by combining 2D COSY and TOCSY and 3D HSQC-TOCSY information. The chemical shifts of highly conserved kringle structure reporter groups, including the upfield-shifted "fingerprint" Leu46 H^{δ} doublet at −0.966 and 0.214 ppm as well as the downfield-shifted Trp25 $H^{\varepsilon 1}$ (11.556 ppm) and Trp62 $H^{\varepsilon 1}$ (10.490 ppm), are indicative of "native" folding for the recombinant K3 domain (*31*, *50*). Thus, 95.6% (461/482) of ^{1}H resonances and 75.6% (96/127) of ^{15}N resonances of the rK3 construct were accounted for. CcpNmr Analysis 1.0 (*51*) was used for assignment. X-Pro peptidic bonds were found to be *trans* except for the Thr29−Pro30 pair that was *cis* (*52*), in agreement with the reported crystallographic structure of K3 within angiostatin (*22*). $^{3}J_{HNH\alpha}$ values were extracted from the HNHA experiment (*53*) and used to delimit five φ torsion angle constraint categories via the Karplus equation (*54*) with coefficients of 7.09, −1.42, and 1.55 (*55*, *56*). χ_1 angles were estimated on the basis of $^{3}J_{NH\beta}$ and $^{3}J_{HNH\alpha}$ values, respectively extracted from HNHB and COSY experiments (*41*), as well as H^{α}−H^{β} and NH−H^{α} NOEs extracted from the 2D NOESY (τ_{mix} = 100 ms) spectrum (*57*). (The information outlined in the above is summarized in Supporting Information, Figure S1.)

Sequence-specific ^{1}H and ^{15}N NMR assignments of r(K57D)K3 were determined via the same protocol as for rK3; 73.1% (348/476) of ^{1}H resonances and 72.2% (91/126) of ^{15}N resonances of the r(K57D)K3 construct were accounted for.

Structure Calculation. A preliminary structural model for rK3, generated using the SWISS-MODEL server (*58*−*60*) in automated modeling mode, was used to assist in an initial manual assignment of NOESY cross-peaks. In this procedure, the crystal structure of angiostatin was excluded from the list of templates. Close to 40% of all NOEs from 2D NOESY (τ_{mix} = 100 ms) and 3D HSQC-NOESY (τ_{mix} = 100 ms) spectra were unambiguously assigned, with the remainder converted to ambiguous constraints.

Iterative automated assignment of the NOE spectra was carried out via ARIA 2.2 (*61*) integrated with CNS 1.2 (*62*) following a standard protocol (*63*). Representative structures from each run were reimported to CCPN Analysis and used to generate synthetic NOESY cross-peak tables based on the suggested interproton distances, thus interactively improving the number and quality of NOE constraints between runs. While keeping existing NOE assignments frozen, cross-peaks from the 2D NOESY (τ_{mix} = 200 ms) experiment were next picked, assigned and disambiguated in the same interactive manner, and used to augment the number of distance constraints fed to the structure calculation protocol. A total of 48 hydrogen bonds were identified via WHAT IF (*64*, *65*) and verified by direct inspection; when found consistent, they were added as structural constraints (Supporting Information, Table S1).

Disulfide bond constraints were next incorporated in the rK3 structure calculation. Cystine residues are known to cluster into two mirror-image conformations based on the sign of the (central) dihedral angle χ_3 (*66*), with variations in χ_1 and χ_2 values resulting in 10 distinct distribution modes (*67*). Modes

Table 1: NMR Structure Statistics[a]

energies (kcal/mol)	
total	−2837 ± 61
bond	25 ± 1
angle	111 ± 5
improper	229 ± 25
van der Waals	756 ± 27
electrostatic	−2874 ± 55
Ramachandran plots	
PROCHECK	
core regions (%)	65.01 ± 2.47
allowed regions (%)	32.87 ± 2.66
generously allowed regions (%)	0.71 ± 0.86
disallowed regions (%)	1.40 ± 0.00
MolProbity	
favored (98%) regions (%)	78.6
allowed (>99.8%) regions (%)	98.0
residue properties	
bad contacts	0
bond lengths	1.8 ± 0.2
idealized geometry rmsd (Å)[b]	
bonds (Å)	0.0044 ± 0.0001
angles (deg)	0.56 ± 0.01
improper (deg)	1.44 ± 0.08
averaged structure rmsd (Å)[c]	
pairwise	
backbone (N, CA, C′)	0.30 ± 0.06
all heavy atoms	0.70 ± 0.10
to the average structure	
backbone (N, CA, C′)	0.35 ± 0.07
all heavy atoms	0.75 ± 0.12

[a]Statistics over the selected bundle of 20 NMR structures. [b]Idealized covalent geometry based on the Aria 2.2 parallhdg5.0 force field (*61*). [c]rmsd values are for the C1−C80 segment (i.e., the kringle domain proper, without tails).

observed in rK3 were weighted vis-à-vis those previously determined in Pgn kringle structures (*24*) and the χ_3 dihedral angles within the core Cys22−Cys63/Cys51−Cys75 disulfide pair included as additional constraints, decreasing both the total energy by ∼5% and the backbone heavy atom average root-mean-square deviation (rmsd) for the structural bundle. A final iteration in explicit water, while marginally affecting the side chain definition, led to considerable improvement in the backbone heavy atom rmsd from 0.65 to 0.35 Å.

A total of 2176 final interproton distance constraints were estimated from the NOE buildup of unambiguously assigned cross-peaks, of which 1248 were determined via AQUA (*68*−*70*) to be conformationally restricting and used to compute the structure of rK3. The latter include 350 intraresidue ($i \rightarrow i$), 287 sequential ($|i-j| = 1$), 182 medium-range ($1 < |i-j| < 5$), and 429 long-range ($|i-j| \geq 5$) distance constraints (Supporting Information, Table S1).

An ensemble of 20 conformers was selected on the basis of total energy value, its quality verified by stereochemical analysis with PROCHECK-NMR 3.5.4 (*68*) and MolProbity 3.15 (*71*) and the pairwise rmsd of the conformers calculated using MOLMOL 2K.2 (*72*) (Table 1). Each conformer in the bundle was analyzed for secondary structure via XTLsstr 1.3.9 (*73*) and STRIDE 1.0 (*74*), followed by visual inspection. Thus, as expected, the least divergence between structures in the bundle is observed for stretches 1−4, 13−30, 48−54, 60−66, and 71−77 of residues that are constrained by a large number of NOEs. In particular, residues in the vicinity of the core cystine pair Cys22−Cys63/Cys51−Cys75 are well localized and happen to essentially coincide with those in the X-ray crystallographic structure. The representative structure, defined as the one closest to the mean, was further investigated using WHAT IF (*64*, *65*): it forms the basis for all subsequent discussions. An electrostatic surface map of rK3 was generated using APBS (*75*). Molecular representations were prepared with PyMOL 1.1 (*76*).

Structural Modeling of K3 Homologues. Two sets of homologous kringle folds proper, excluding N-terminal Thr-Tyr-Gln and C-terminal Asp-Ser "tail" segments, were generated via molecular modeling and referred to as "K3" constructs. The first consists of mutated hPgn rK3 structures and exploits as templates either the NMR rK3 structure or the one extracted from the angiostatin crystallographic structure (*22*). Mutations include K57D, corresponding to the r(K57D)K3 construct used in our NMR experiments, K57E, K57A, K57Q, K57S, and H64Y, as well as the double mutation K57E/H64Y. Models were generated via PyMOL (*76*), followed by 500 steps of steepest descent energy minimization of the mutated residue(s) and their immediate positional neighbors *in vacuo*, as specified by the GROMOS96 43B1 parameters set embedded in SWISS-PdbViewer (*59*). The second set consisted of orthologous and paralogous wild-type Pgn K3 folds of other species (Scheme 2). The latter models were generated by threaded sequence homology to the NMR structure of rK3 using SWISS-MODEL in project mode (*58*−*60*, *77*).

Ligand Titration Experiments. NMR ligand titration experiments on r(K57D)K3 were carried out at 25 °C in 9:1 H_2O:D_2O, pH* 7.0. Since the sequential assignments for rK3 had been obtained at pH* 5.7, a series of $^1H/^{15}N$ HSQC spectra were recorded on r(K57D)K3 within the pH 5.7−7.0 range in order to verify identity of cross-peaks at pH* 7.0 (Figure 1). The experiments did not uncover any significant NH shifts except for intrinsically pH titratable side chain groups, indicating that the overall structure of the kringle remains essentially unchanged within the pH range (not shown). However, some resonances, notably the Arg36 amide NH group, become undetectable at neutral pH.

Kringle−ligand equilibrium association constants (K_a) for AMCHA were assessed by monitoring the chemical shift changes following incremental additions of ligand. A 40 mM stock solution of AMCHA was added in small aliquots (2−100 μL) to 475 μL of a 0.28 mM solution of ^{15}N-r(K57D)K3 followed by the recording of $^1H/^{15}N$ HSQC spectra after each addition. The progressive chemical shift change of amide cross-peaks (Figure 5, inset a) was assessed separately in the ^1H and ^{15}N dimensions throughout the titration series. The amplitudes of net chemical shift perturbations ($\Delta\delta = |\delta_{bound} - \delta_{free}|$) of all backbone amide cross-peaks were expressed as the normalized quadratically weighed perturbations in both dimensions:

$$\Delta\delta = \sqrt{\frac{(\Delta\delta_{1H} \times 600.333 \times 10^6)^2 + (\Delta\delta_{15N} \times 60.838 \times 10^6)^2}{(600.333 \times 10^6)^2 + (60.838 \times 10^6)^2}} \quad (1)$$

and used to map the AMCHA binding site. Using these data, the fraction of bound kringle (Δp) was determined after each ligand addition (*78*) from selected resonances' chemical shifts (δ_{obs}):

$$\Delta p = \frac{\delta_{obs} - \delta_{free}}{\delta_{bound} - \delta_{free}} \quad (2)$$

Unspecific interactions that resulted in linear baseline trends were least-squares fitted and subtracted from the experimental data. The affinity toward the ligand was thereafter determined from the titration profiles of individual residues by

iterative fitting of K_a and the normalization parameter P to the hyperbolic Langmuir adsorption isotherm (*29*) until convergence ($P = 1$) (*16*):

$$\Delta p = \frac{PK_a[S]}{1 + K_a[S]} \quad (3)$$

where [S] stands for the free ligand concentration determined after each addition of ligand. Origin 6.1 (Originlabs, Northampton) was used for the nonlinear least-squares fits. After each iteration step, data outliers were identified and discarded by assuming a normal distribution and applying the three-sigma exclusion criterion. The reported K_a values and errors are weighted as the square of the normalized inversed uncertainties associated with each measurement (*79*).

Ligand Docking. Models of bound zwitterionic ligands to K3 were simulated via AutoDock 4.0 (*80*−*82*). AutoDock has been proven to afford a robust tool to visualize binding interactions between K5 and AMCHA as well as fatty acids (*16*). The approach predicts an ensemble of flexible ligand conformations posed at the corresponding binding sites on macromolecular receptor targets; it involves a random Lamarckian genetic algorithm search combined with energy minimization. The bound state conformations are scored according to a free energy cost function (ΔG^*) that accounts for Lennard-Jones and Coulomb electrostatic interactions, directional hydrogen bonding, an entropic contribution based on the loss of ligand conformational degrees of freedom, and a desolvation term.

K3 models were prepared for docking by regularizing the PDB files with MolProbity (*71*) and WHAT IF (*64*) followed by addition of atomic charges, calculated using the PARSE force field (*83*) at pH 7.0 in water with the program PDB2PQR (*84*) augmented by PROPKA (*85*). The ligand molecules were modeled as zwitterions with ACD/ChemSketch 12.01 (Advanced Chemistry Development, Toronto, Canada) and then prepared for AutoDock using AutoDockTools 1.5.2 (*86*) by merging nonpolar hydrogens and defining torsion trees. The aliphatic ligands 6-aminohexanoic acid (6-AHA) and N^{α}-acetyl-L-lysine (AcLys), as well as *p*-benzylaminesulfonic acid (BASA), an aromatic ligand, were generated in this fashion. In the case of BASA, the resulting partial charges were distributed evenly over all ligand atoms.

AutoDock 4.0 is unsuitable to manage directly the ring flexibility associated with cyclic molecules (*87*). The aliphatic cyclic ligand AMCHA in both chair- and twist-boat conformations was tested on the crystallographic structure of hPgn K4, a kringle with a high affinity for AMCHA ($K_a \sim 159 \pm 2$ mM^{-1}) (*29*). Improvement followed from using the energetically favored chair conformation, in line with the solution structure of the hPgn K2−AMCHA complex (*11*); hence, all AMCHA docking experiments on (K57D)K3 were carried out with the ligand in the chair conformation.

The docking space was defined as a grid of 126 × 126 × 126 points with a spacing of 0.375 Å: this yields a cube with an edge length of 47.3 Å that encompasses the entire kringle domain. A total of 1280 docking runs were performed for each ligand, whereby genetic algorithm operations were applied to a population of 300 individuals as delimited by a maximum of 27000 generations/2500000 energy evaluations. Additional long-term dockings (maximum of 25000000 energy evaluations) were carried out while keeping selected kringle side chains flexible and restricting the docking space by using a spacing of 0.188 Å on the same grid: this corresponds to a cube with an edge length of 23.7 Å centered at the binding site. Parameters identifying mutation, crossover, and elitism weights were left at their default values of 0.02, 0.80, and 1, respectively. Docking results were clustered using an rmsd of typically 2.0 Å and ranked according to the change in free energy (ΔG^*) estimated by AutoDock.

The docked conformation associated with the lowest ΔG^* value at the locus of the binding site was selected; additionally, we report $\langle \Delta G^* \rangle$, the corresponding cluster ensemble average ΔG^*. This enabled for defining an effective association constant (K_a^*) for the interaction:

$$K_a^* \equiv e^{-\langle \Delta G^* \rangle / RT} \quad (4)$$

where R is the ideal gas constant and T, the thermodynamic temperature, is set to 298.15 K. Ligand−kringle pairs that did not yield dockings at the putative LBS locus were assigned $K_a^* = 0$. For a group of ligands an average association constant ($\overline{K_a^*}$) was defined, weighted according to the normalized inverse uncertainties squared.

Phylogenetic Analysis. A phylogenetic analysis of the K3 domain was carried out via MEGA4 (*88*) using the amino acid sequences from Scheme 2, aligned with ClustalW (*89*), which encompass 13 orthologous and 38 paralogous Pgn K3 domains. A rooted tree was generated based on the proportion of amino acid difference using the unweighted pair-group method with arithmetic mean (UPGMA) (*90*) and its accuracy tested by bootstrap analysis (*91*). The consensus tree, inferred from 100000 replicates, is taken to represent the evolutionary history of the 51 taxa analyzed. Evolutionary distances were computed using the Poisson correction method (*92*). Positions containing missing data were eliminated only during pairwise sequence comparisons, retaining 78 amino acid positions in the data set.

RESULTS AND DISCUSSION

Solution Structure of Kringle 3. A stereoview of the superposed NMR folds in the ensemble of 20 final K3 structures is shown in Figure 2A. Residues within the domain proper (i.e., excluding tails) are well-defined, with a pairwise rmsd of 0.30 ± 0.06 Å for superposed backbone atoms and 0.70 ± 0.10 Å for all heavy atoms. The closest to average structure (taken as the representative NMR structure) has dimensions of ∼31.9 Å × 38.0 Å × 23.0 Å: it differs from the one within the angiostatin crystallographic structure (*22*) by a backbone heavy atom rmsd of 1.13 Å (Figure 2B). No distance violations >0.5 Å or torsion angle violations >5° were observed.

Stereochemical analysis of the representative structure using MolProbity (*71*) indicates that all residues fall in the energy-allowed regions of a Ramachandran plot, 76.5% appearing in the favored zones, and yielded a MolProbity score, a weighted sum of clashes, Ramachandran not-favored and rotamer outliers (*71*), of 3.33. This result is comparable to the solution structures of K2 (90.1% of residues in allowed regions/76.5% of residues in favored regions/MolProbity score of 3.35) (*11*) and of K5 (100%/85.6%/3.58) (*16*).

The core region populations in the K3 Ramachandran steric energy plot indicate occurrence of β-strands and β-turns as well as right-handed (α and 3_{10}) and left-handed (3_1) helices (*93*). Identified elements of secondary structure, when averaged over the bundle of 20 structures, are ∼19.1 ± 1.8% β-turn, ∼8.4 ± 2.0% 3_1-helix, ∼8.2 ± 1.8% β-strand, ∼3.3 ± 2.3% 3_{10}-helix, and ∼2.8 ± 1.6% α-helix. Hydrogen bonds connect two short

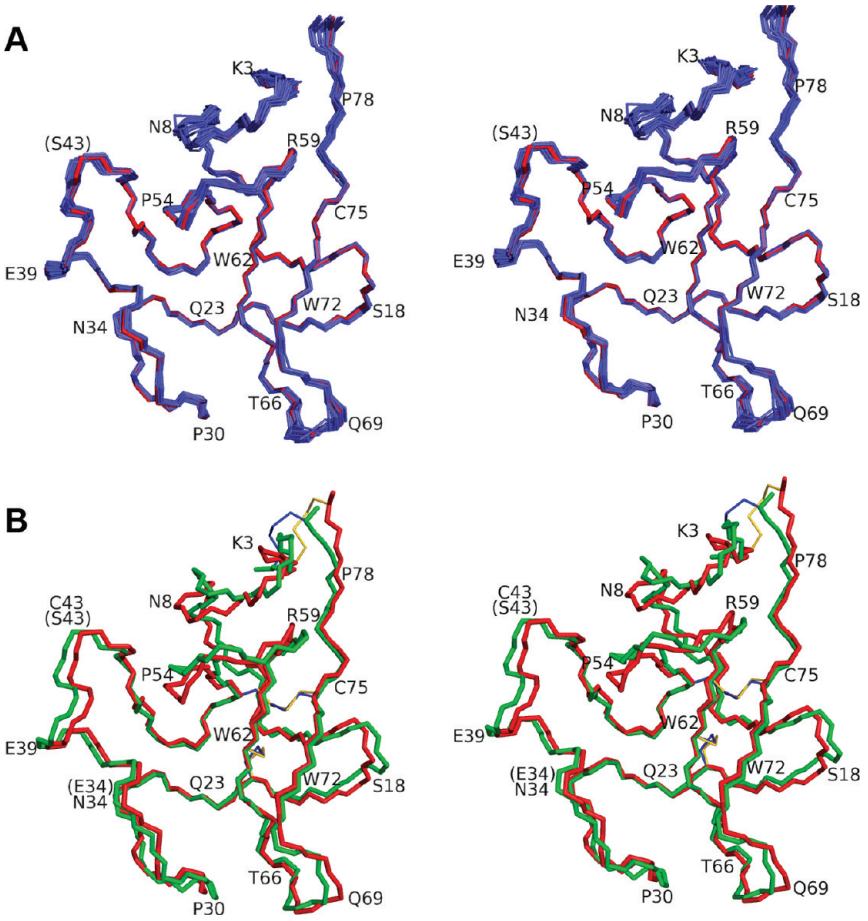


FIGURE 2: Kringle 3 three-dimensional structures: stereoviews of the Cys1−Cys80 polypeptide backbone. Selected residues are labeled; flexible tail segments preceding Cys1 and following Cys80 are omitted. Parentheses denote loci where residues have been mutated from the wild type to avoid dimerization [Ser43 (*32*)] or glycosylation [Glu34 (*22*)]. (A) Superposition of 20 NMR structures. The closest to the average (representative) structure is shown in red trace: mean backbone rmsd = 0.35 ± 0.07 Å. (B) Closest to average NMR structure (red) superposed to the X-ray (angiostatin) crystal structure (green) (*22*). The backbone heavy atom rmsd between the two structures is 1.13 Å. Disulfide bridges are colored yellow in the NMR structure and blue in the crystal structure: the Cys22−Cys63 and Cys51−Cys75 bridges for the two structures essentially coincide.

antiparallel β-sheets in quasi-orthogonal configuration (Figure 3A): the first is defined by the paired strands Trp62−Thr65/Trp72−Tyr74, stabilized by Cys63-$H^N\cdots$O=C′-Glu73, Thr65-$H^N\cdots$O=C′-Arg71, Glu73-$H^N\cdots$O=C′-Cys63, and Cys75-$H^N\cdots$O=C′-Pro61 backbone−backbone hydrogen bonds; the second is composed by pairs of short strands Ala14−Thr16/Thr21−Gln23, stabilized by Thr16-$H^N\cdots$O=C′-His20 and Cys22-$H^N\cdots$O=C′-Ala14 backbone−backbone hydrogen bonds. Other conspicuous hydrogen bonds include Val17-$H^N\cdots$O=C′-Tyr74 and Gln23-$H^N\cdots$O=C′-His64, common to all hPgn kringles (*8*−*16*, *22*), as well as Ser18-$OH^\gamma\cdots$O=C^ε-Glu73 originally observed in the K2 solution structure (*11*). Furthermore, K3 includes two consecutive but distinct helical elements (Figure 3A): an α-helical turn (Thr37−Asn40, characterized by a C′=O (i)$\cdots$NH (i+4) hydrogen bond and $\varphi = -62.6°$) that resolves into a 3_{10}-helical turn (Phe41−Lys44, characterized by a C′=O (i)$\cdots$NH (i+3) hydrogen bond and $\varphi = -67°$, $\psi = -31°$), of common occurrence for short 3_{10}-helices (*66*). Five poly(Pro)II 3_1-helices, which lack intrahelix hydrogen bonds, are also identified (Figure 3B): the first is located at the N-terminus (Gln(−1)−Leu2); the second extends from Thr21 to Cys24 and encompasses the Cys22−Gln23 β-strand discussed above. Additional 3_1-helices are formed by segments His31−His33 and Tyr50−Arg52; furthermore, a longer 3_1-helix extends from Trp72−Pro78 near the C-terminus, overlapping with the Trp72 -Tyr74 β-strand identified earlier.

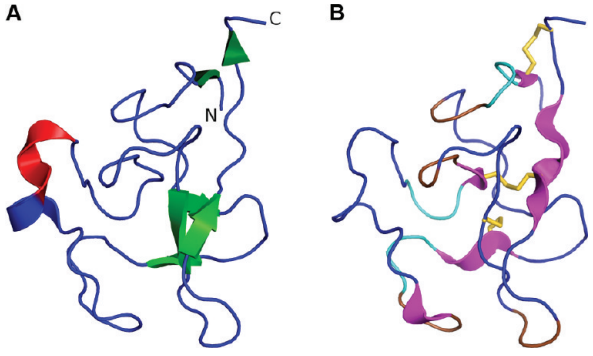


FIGURE 3: Kringle 3 NMR fold. Secondary structure elements are displayed in ribbon representation: (A) α-helical turn (blue), 3_{10}-helical turn (red), β-strands and β-bridge (green); (B) 3_1-helix (pink), β-turn (cyan), β-turn without a hydrogen bond (brown), disulfide bonds (yellow trace). N- and C-termini are indicated.

The two core cystines (Cys22−Cys63/Cys51−Cys75), which span disulfide bridges within ~4 Å and slanted ~30°, thus appear ensconced by secondary structure elements. The 3_1-helical motifs cluster around the cystine links, the main contribution to such structure stemming from the stretches 21−24 and 72−78. Hence, both length and location of 3_1-helical segments match earlier observations (*11*), a structural signature of kringle domains not mentioned in the crystallographic analyses (*8*, *10*, *12*, *14*, *15*, *22*).

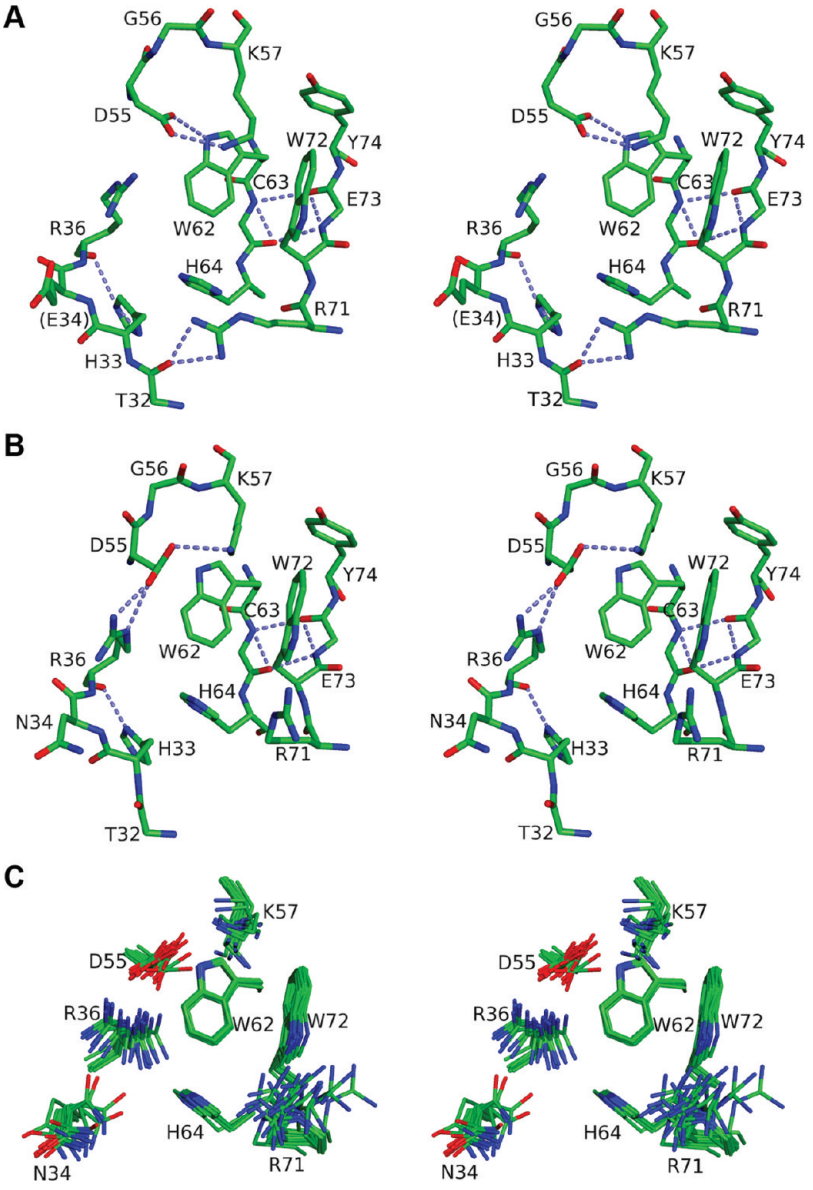


FIGURE 4: Kringle 3 structures: stick representation stereoviews of the canonical binding site. All panels are shown in the same orientation; carbon, nitrogen, and oxygen atoms are indicated in green, blue, and red, respectively. (A) X-ray crystal structure (*22*): Glu34, mutated from Asn to prevent glycosylation, is labeled in parentheses. (B) Closest to the mean NMR structure. Segments His33−Arg36, Asp55−Asp57, Trp62−His64, and Arg71−Tyr74 are displayed; suggested interresidue hydrogen bonds are dotted (cyan). For clarity, only polypeptide backbone atoms are shown for residues Cys63 and Glu73. (C) Superposition of selected side chains from 20 NMR structures.

A total of seven β-turns are suggested on the basis of $d_{(i)-(i+3)}$ backbone distance constraints, three of which are stabilized by a C′═O (i)···HN (i+3) hydrogen bond: Lys3−Thr5, Trp25−Ala27, and Asp47−Asn49, denoted by a cyan trace in Figure 3B. The analysis illustrates the strong ordering effect of secondary structure on protein structures, as all β-strands and most helical structures exhibit rmsd minima. An exception is the highly disordered Thr37−Lys44 loop exposed at the protein surface. In contrast, the non-H-bonded β-turns shown in Figure 3B (brown trace) are readily identified as being among the least defined stretches (compare with Figure 2A).

The Ligand Binding Site. The NMR and X-ray structures differ in the hydrogen-bonding pattern between the Asp55 side chain and residues at the binding site. Thus, the Asp55 χ_1 value that corresponds to *gauche* (+) in the crystal becomes *trans* in solution. As a result, the hydrogen bond Asp55-$COO^{\delta 1}\cdots HN^{\varepsilon 1}$-Trp62 (Figure 4A) is replaced by Asp55-$COO^{\delta 1}\cdots HN^{\varepsilon}$-Arg36 in solution (Figure 4B). This discrepancy may stem from experimental conditions: while the NMR structure was obtained for the protein in the absence of ligand, the X-ray model of angiostatin (K1−K2−K3) shows a molecule of bicine buffer parked within the K3 binding site groove (*22*). It was noticed that, as determined by MolProbity (*71*), rotamers for both the Asp55 and Lys57 side chains in the solution structure are energetically preferred over the crystallographic counterparts, occurring in 17.2%/74.7% of cases in the Top 500 angle database[4] for the former versus 10.0%/41.3% for the latter.

Interestingly, the network of hydrogen bonds within the NMR structure is similar to the pattern previously observed in K4-type kringles of human Apo(a) (*94*) where Arg36 is a strictly conserved residue (*95*). In the case of human Apo(a), a hydrogen bond involving Arg36 and Asp55 was suggested to weaken the kringle's lysine-binding affinity by restricting the mobility of the LBS

[4]Top 500 Database, http://kinemage.biochem.duke.edu/databases/top500.php.

cationic and anionic centers (*94*). Arg36 was identified as a key residue in human Apo(a) K4 repeats and may thus be similarly important in stabilizing the structure of hPgn K3. Overall, key side chains at the canonical binding site are generally well-defined in the solution structure (Figure 4C). An exception is the largely solvent-exposed Arg71 side chain, which translates into local structure disorder within the rK3 NMR bundle computed for the kringle in explicit water (Figure 4C). In contrast, the crystallographic angiostatin K3 structure shows this side chain in a buried location that leads to the occurrence of two Thr32-CO$\cdots$N$^{\eta 1}$/N$^{\eta 2}$-Arg71 backbone−side chain hydrogen bonds (Figure 4A).

The canonical binding site of wild-type K3 manifests itself as an exposed hydrophobic surface that conforms a shallow groove, relatively widened circa the conserved Arg71 cationic center (Figure 4B). Lys57 projects its side chain over the hydrophobic groove, consistent with the known intimate interaction of Lys-type ligand methylene groups with lipophilic side chains that line the LBS (*78*). A minor difference between the X-ray and NMR structures (Figure 4A,B) arises from the Lys57 χ_2 and χ_5 dihedral angles [X-ray: *trans*/*gauche* (+), NMR: *gauche* (−)/*trans*], most likely a reflection of side chain rotameric flexibility. However, in both structures, the net effect is to reduce exposure of the lipophilic surface lined, among others, by the Trp62 and Trp72 side chain indole rings (*23*). Indeed, NOEs between Lys57 H$^{\beta}$, H$^{\gamma}$ and Trp 62, 72 ring protons are observed, supporting such intramolecular hydrophobic interactions.

Ligand Binding to K57D Kringle 3. Knowledge of the solution structure of K3 (Figure 2A) is a prerequisite for the analyses of both *in vitro* and *in silico* ligand-binding experiments. Superposition of the ^{1}H/^{15}N HSQC spectra of uniformly ^{15}N-labeled rK3 and r(K57D)K3 (Figure 1) reveals that, except for the expected residue 57 signals (denoted by red cross-peak labels), the chemical shift pattern remains only slightly perturbed on going from the wild-type to the mutated protein. Hence, the K57D replacement would appear not to affect the fold of the K3 domain. On this basis, structures for the hPgn (K57D)K3 mutant were generated via homology modeling starting from the solution and crystallographic structures of the wild-type domain.

Upon addition of AMCHA to a solution of ^{15}N-r(K57D)K3, a number of ^{1}H/^{15}N HSQC cross-peaks shift (Figure 5A). In line with similar studies on kringle modules with canonical (i.e., active) LBS (*8*, *11*, *14*, *29*), addition of the zwitterionic ligand to r(K57D)K3 shifts the most polypeptide chain backbone amide NH resonances arising from residues Asp55 ($\Delta\delta \sim 0.449$ ppm), His64 ($\Delta\delta \sim 0.420$ ppm), and Trp72 ($\Delta\delta \sim 0.462$ ppm), i.e., all key residues of the canonical LBS (Figure 5B). Interestingly, NH resonances of both Asn34 and Arg71 respond to AMCHA presence: $\Delta\delta \sim 0.325$ ppm and $\Delta\delta \sim 0.105$ ppm, respectively. As to Arg36, the amide NH signal, apparent at pH* 5.7, was undetectable at pH* 7.0 and throughout the ligand titration experiment (Figure 5), as reported for the same residue in kringles of human Apo(a) (*94*, *96*, *97*). Indeed, the perturbation of Asn34, concomitant with that of Thr37 ($\Delta\delta \sim 0.261$ ppm) may echo the response to ligand of the neighboring Arg36.[5] Other significantly perturbed resonances belong to residues that are either part (Trp62), or immediate neighbors (Tyr74), of the canonical LBS. Surprisingly, Asn45 and Tyr50 NH groups, in the fold relatively remote from the LBS, also result perturbed. We speculate that, while these residues may weakly interact with AMCHA, they are also likely to echo perturbations elsewhere in the protein or be sensitive to interkringle interaction(s) which break upon addition of ligand.

K_a values were determined by nonlinearly fitting ligand titration profiles (Figure 5, inset b) for individual residues to eq 3. Three independent criteria were considered: the amplitude of the chemical shift perturbation, the quality of the nonlinear fit to Langmuir isotherms, and presence of the residue in the putative binding site. On this basis, we derive $K_a \sim 5.232 \pm 0.026$ mM^{-1} from the weighted average of 16 fits. This value is close to that reported for the binding of AMCHA to K2 [$K_a \sim 7.3 \pm 0.6$ mM^{-1} (*29*)]. Interestingly, if the K_a is estimated by fitting signals stemming from the introduced Asp57 residue solely, it yields $K_a \sim 12.56 \pm 0.29$ mM^{-1}, a value that is closer to 12.66 ± 0.46 mM^{-1} estimated for the r(K57D)K3−AMCHA complex via fluorometric ligand titration (*32*) and Scatchard plot fit (*98*). From the data in Figure 5B, AMCHA-induced ^{1}H/^{15}N HSQC perturbations mapped on the (K57D)K3 molecular surface (Figure 6A) show a strong correlation with the locus of the canonical kringle LBS (*23*).

Zwitterionic Ligand Docking to K57D Kringle 3. The NMR AMCHA titration experiment confirms results by fluorescence quenching (*32*) that the K57D mutation introduces affinity of K3 for AMCHA and other zwitterionic Lys mimics. On the basis of the model (Figure 4) and assuming the mutant exhibits no significant conformational change, this ought to result in a gain in exposure of the putative binding site since, relative to a Lys residue, the Asp side chain is sterically less obtrusive.

AMCHA was docked to the (K57D)K3 structure via AutoDock 4.0. It is gratifying to find that the predicted pose of the ligand (Figure 6C) correlates well with the ligand-induced ^{1}H/^{15}N HSQC perturbation map (Figure 6A). Furthermore, the LBS dipolar electrostatic configuration optimally matches the bound zwitterionic ligand dipole length $\ell \sim 6.6$ Å (Figure 6D). Revealingly, the model (Figure 6C,D) shows the docked ligand "sandwiched" between the quasi-parallel planes defined by the Arg36 guanidino and the Trp72 indole groups.

The above results prompted *in silico* exploration for binding of three additional lysine mimics known to interact with kringle domains (*29*) (Figure 7A): 6-AHA, a linear, aliphatic ω-aminocarboxylic acid ($\ell \sim 7.5$ Å); AcLys, a C-terminal Lys analogue ($\ell \sim 7.2$ Å); and BASA, a para-sulfonated aromatic amine ($\ell \sim 6.6$ Å). Among these, 6-AHA and BASA had been shown to interact with (K57D)K3 by the fluorescence quenching studies (*32*). Besides the structure homology-modeled from the rK3 NMR structure discussed above, a second (K57D)K3 structure was modeled starting from the K3 X-ray crystallographic structure: its sequence differs from that of the NMR-based model in that they carry "inert" mutations N34E and C43S, respectively.

Reassuringly, AutoDock predicts all four ligands to dock at the engineered LBS, and except for small qualitative differences in the conformation of the docked ligands discussed below, the ligand−kringle complex models derived from the NMR (Figure 7B) and X-ray (Figure 7C) structures are closely similar. Thus, the cationic amino groups are predicted to consistently interact with the anionic fulcrum formed by the γ-carboxylate groups of Asp55 and Asp57. Notably, the distance separating the ligand N$^{\varepsilon}$ atom from Asp57 C$^{\gamma}$ (3.0 ± 0.2 Å, average among the lowest free energy docked conformations of all four ligands) is marginally lower than the corresponding distance to

[5]There is a deletion at site 35 (Scheme 1).

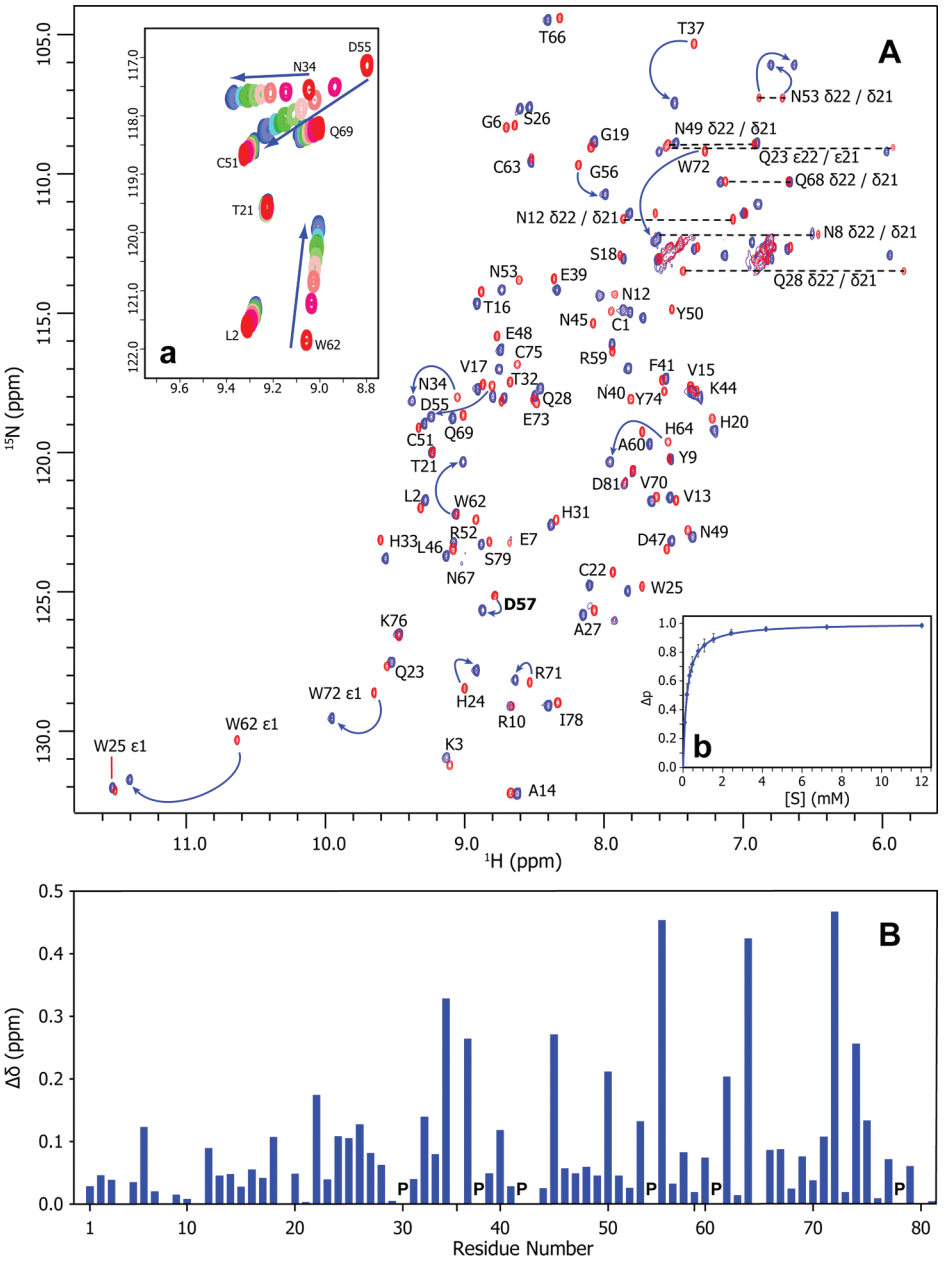


FIGURE 5: AMCHA binding to K57D kringle 3: NMR response. (A) ^{1}H/^{15}N HSQC spectra of (K57D)K3 in the absence (red) and presence (blue) of AMCHA (ligand:protein 62:1). Most pronounced ligand-induced cross-peak shifts are indicated by arrows. Signals arising from Asn $H^{\delta 21}/H^{\delta 22}$ and Gln $H^{\varepsilon 21}/H^{\varepsilon 22}$ geminal proton pairs are connected by dashed lines. Inset a: Superposed spectra recorded after each addition of AMCHA, illustrating the progressive response (blue arrows) of select resonances. Inset b: Ligand titration profile. Δp represents the weighted average fraction of ligand-bound protein versus the free ligand concentration [S]. Langmuir adsorption isotherm curves were obtained by iterative nonlinear fit of the data, each data point representing an average of 16 perturbed resonances; error bars reflect the range of values. The continuous curve was generated by fitting the weighted average of all calculated K_a values. Data were recorded at 600 MHz (^{1}H), 25 °C, and pH* 7.0. (B) Average ^{1}H, ^{15}N backbone amide resonance shifts induced by ligand binding ($\Delta\delta$, see text) against sequence number are represented by bars. Gaps denote Pro (P) and amide NH groups that do not yield detectable signals at pH* 7.0. Sites 36 and 59, following deletions relative to the K5 sequence, are indicated by ticks on the abscissa axis.

Asp55 C^{γ} (3.2 ± 0.1 Å), a slight preference that is in line with reports for other hPgn kringles (*11*, *99*). Interestingly, as suggested by the AutoDock simulations, none of the tested ligands docks at the back side of the kringle domain (not shown), supporting our contention that the observed NMR perturbation of Asn45 and Tyr50 NH resonances in the presence of AMCHA (Figure 6B) is unlikely to arise from an affinity of these residues for the dipolar ligands.

As a negative control, AutoDock runs on the above ligands were implemented on the solution and crystal structures of *wild-type* human Pgn K3. Out of all the calculated conformations for each of the ligands (1280 × 4), *none* docked to the crystal structure wild-type binding site (Supporting Information, Figure S2). Docking to the solution structure yielded slightly more nuanced results: while the cyclic ligands AMCHA and BASA show no affinity for the K3 binding site, marginal dockings of 6-AHA and AcLys were predicted (frequencies of 2.7% and 3.3%, respectively), albeit with low affinity. Moreover, the ligands docked in unorthodox conformations (Supporting Information, Figure S2).

Free Energies of Docking. On the basis of the AutoDock binding free energy estimates, it is hinted that the aromatic

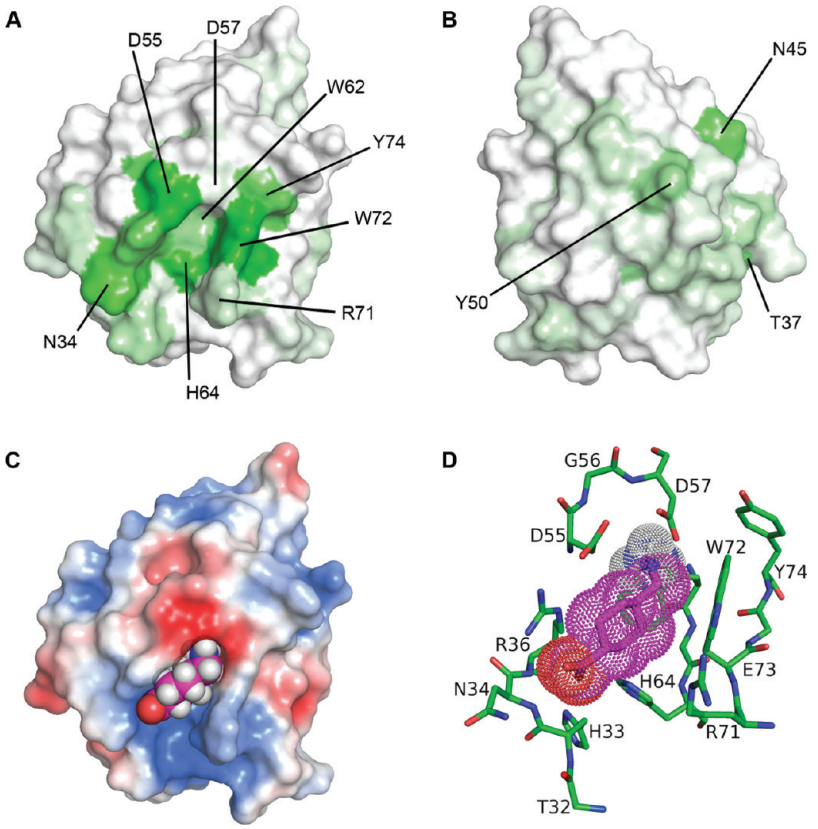


FIGURE 6: AMCHA docking to the engineered lysine-binding site in K57D kringle 3. The Cys1−Cys80 contact surface is shown. (A) Front view; (B) back view: color intensity (green) is proportional to the composite ^{1}H, ^{15}N backbone amide chemical shift changes induced by ligand binding (see text). Unobserved shifts are denoted by averaging values of the two (preceding and following) sequential residues. (C) AMCHA docked on the electrostatic surface of (K57D) (front view). Electrostatic potential was calculated on the basis of the PARSE force field (*83*) at 298.15 K using the APBS 1.1 plug-in for PyMOL (*75*) and colored for the solvent-accessible surface from −3 kT/e (red) to +3 kT/e (blue). The ligand is shown in effective van der Waals radii, space-filling representation. The pose of the ligand was predicted via AutoDock 4.0. (D) Expansion of the binding site with docked ligand (stick representation). Residue segments His33−Arg36, Asp55−Asp57, Trp62−His64, and Arg71−Tyr74 of the engineered K57D binding site are displayed. Only backbone atoms are shown for Glu73. Ligand atoms are in s tick representation and superposed with van der Waals surfaces (dotted); only polar hydrogen atoms are shown. The view matches the orientation of the wild-type binding site shown in Figure 4. In (C) and (D), the ligand atoms are denoted in magenta (carbon), blue (nitrogen), and red (oxygen).

effector BASA is the favored ligand for (K57D)K3 ($\langle\Delta G^*\rangle$ = −6.62 ± 0.06 kcal/mol) followed by the cyclic aliphatic AMCHA ($\langle\Delta G^*\rangle$ = −6.21 ± 0.11 kcal/mol). Furthermore, it is suggested that AcLys and 6-AHA exhibit the lesser affinity among the four investigated ligands (Table 2). The *in silico* results thus concur with the trend of the experimentally determined K_a values, supporting the hypothesis (*32*) that the extended surface of the cyclic ligands results in enhanced interaction with hydrophobic side chains of either Trp72 alone or Trp62 + Trp72 jointly in the LBS groove.

Differences between docked structures based on the NMR (solution) and X-ray (crystal) (K57D)K3 models are most notable in the dockings of 6-AHA and AMCHA and may be attributed to the orientation of the Arg71 side chain in the two rigid structures: whereas in the NMR-derived model the extruding Arg71 side chain tends to push the ligand backbone toward the Asp55, Arg36, Asn34 triad (Figure 7B), in the X-ray structure derived model Arg71 adopts a less extended conformation that enables it to interact with the ligand carboxylate group (Figure 7C). Thus, for all the investigated ligands, the γ-carboxylate group is positioned closest to the Arg36 N^{η} atoms (3.5 ± 0.4 Å) in the NMR-derived and to the Arg71 N^{η} atoms (3.3 ± 0.2 Å) in the crystal-derived models. In human Pgn, K4 is the only other kringle domain containing similarly basic residues at positions 34 and 36 (Scheme 1A). For the latter domain, the X-ray crystallographic structure places the corresponding Lys36 residue in close proximity to the ligand carboxylate group (*14*) while Arg71 affords the ligand-preferred cationic center in solution (*100*).

Binding Site Flexibility Effects. In the bound state, the AMCHA N^{ε} atom positions itself essentially equidistant of the Asp55 and Asp57 carboxylate groups, at ∼3.2 and ∼3.3 Å, respectively. In the canonical lysine-binding kringles of hPgn, i.e., K1, K2, K4, the zwitterionic ligand is stabilized by ionic interactions between the ligand carboxylate group and the guanidino group of Arg71 (*10*, *99*). Our rigid kringle docking model of the (K57D)K3 structure however shows AMCHA preferentially interacting with Arg36 $N^{\eta 1}$ (∼3.3 Å), with His64 $N^{\varepsilon 1}$ and Asn34 $N^{\delta 2}$ in close proximity (∼4.5 and ∼4.9 Å, respectively).

Differences between the solution and crystal K3 binding site models may be interpreted to represent snapshots of conformations that arise from the intrinsic flexibility of the exposed Arg71 side chain. Indeed, residues 34, 36, and 71 are relatively disordered in the NMR structural bundle (Figure 4C). In order to determine the preferred ligand anionic group acceptor in the (K57D)K3 domain LBS, a long-term AMCHA docking experiment (see Experimental Procedures) was carried out during

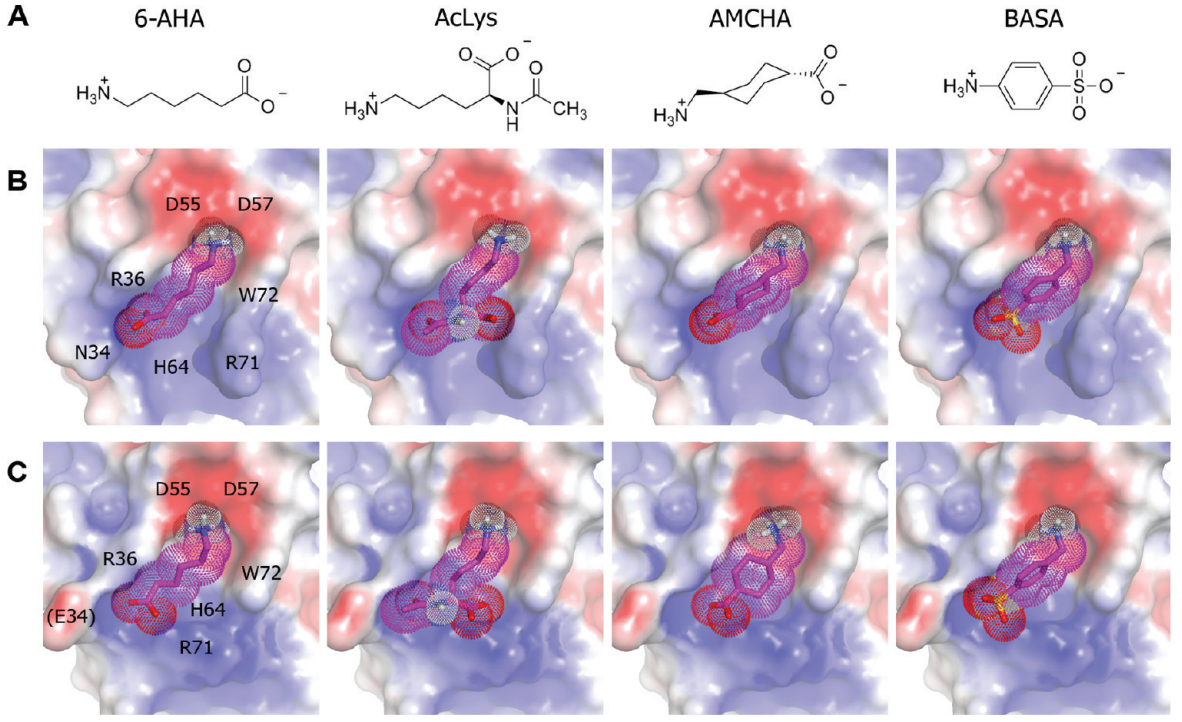


FIGURE 7: Zwitterionic ligands docked to the K57D kringle 3 binding site. Docking was via AutoDock 4.0 on a rigid protein structure; the lowest energy conformations of the docked ligands are shown. Ligand atoms are in stick representation, superposed with Van der Waals surfaces (dotted); carbon, nitrogen, oxygen and sulfur atoms are denoted in magenta, blue, red, and yellow, respectively; only polar group hydrogen atoms are shown. (A) Chemical structures of the ligands. (B) and (C) Protein surface electrostatics (calculated and colored as in Figure 6C) were modeled for the NMR structure of K3 and the X-ray structure of the angiostatin K3, respectively. Selected residues are labeled; residue Glu34, indicated in parentheses, was mutated from Asn to prevent glycosylation (*22*).

Table 2: Zwitterionic Ligand Free Energies of Binding ($\langle \Delta G^* \rangle$) [a] for K57D Kringle 3

data origin	6-AHA	AcLys	AMCHA	BASA
NMR model	-5.64 ± 0.30	-5.54 ± 0.44	-6.21 ± 0.11	-6.62 ± 0.06
X-ray model	-6.45 ± 0.22	-6.55 ± 0.49	-7.21 ± 0.12	-7.52 ± 0.17
experimental[b]	-4.94 ± 0.02	n/a	-5.08 ± 0.01	-5.55 ± 0.02

[a]Dockings were carried out on rigid receptors. $\langle \Delta G^* \rangle$ represents the average of ΔG^* estimates computed via Autodock for the cluster of ligand dockings at the binding site. [b]Experimental ΔG were calculated from K_a values in units of mM^{-1} estimated via fluorometric ligand titrations (*32*). The AMCHA ΔG was calculated using the weighted average K_a between the fluorometric and NMR ligand titrations.

which side chains of Asn34, Arg36, Asp55, Asp57, Arg71, and Trp72 were allowed to rotate freely.

The distance between AMCHA carboxylate and Arg36 + Arg71 guanidino groups, and AMCHA amino and Asp55 + Asp57 carboxylate groups, was used as a filtering criterion: docked conformations in which the ligand came to be positioned > 3.5 Å [6] from both the cationic and anionic centers of the LBS were excluded. It was found that ~30.8% of the dockings concur with the results from the rigid protein docking, where the ligand interacts preferentially with Arg36 with $\langle \Delta G^* \rangle = -6.62 \pm 1.04$ kcal/mol (Figure 8, panel A1). In a further ~6.7% of the cases ($\langle \Delta G^* \rangle = -7.70 \pm 0.84$ kcal/mol), the ligand carboxylate group is within hydrogen bond-forming distance to both Arg36 and Arg71 (Figure 8, panel A2). In most cases (~62.5% of dockings, $\langle \Delta G^* \rangle = -7.18 \pm 1.05$ kcal/mol), the AMCHA carboxylate group interacts exclusively with the Arg71 guanidino group, consistently "tracking" the latter's mobile side chain. (Figure 8, panel A3).

Measured distances between the ligand polar groups and the flexible kringle ionic centers were averaged for each cluster (Table 3). Interestingly, a preference to interact with Arg36 is concomitant with a slightly stronger interaction with Asp55 over Asp57 (Figure 8, panel A1). In contrast, interaction with Arg71 displaces the ligand toward Trp72 and results in a preference for Asp57, the "canonical" model (Figure 8, panels A2 and A3). The large average distance, in all three observed clusters, between the ligand carboxylate C group and Asn34 $N^{\delta 2}$ (Figure 8, Table 3) precludes the possibility of hydrogen bond formation, therefore decreasing the likelihood of the (unprecedented) interaction with the latter side chain as suggested by the rigid kringle docking (Figure 7B,C).

The docking results for the solution structure (K57D)K3−AMCHA complex with flexible side chains are thus in line with the view (*23*, *100*) that Arg71 affords the main cationic center of the LBS. Nonetheless, Arg36 emerges as a secondary binding partner that could facilitate kringle−zwitterionic ligand interactions. Indeed, by increasing the number of possible docked microstates, Arg36 participation contributes additional entropic stabilization while driving the anionic end of the ligand to bind; alternatively, by raising the probability of interaction it may thus increase the affinity for zwitterionic ligands. Furthermore, the flexible docking underscores the critical role of the Trp72 aromatic side chain as a hydrophobic partner (Figure 8). It is interesting to note that there is a degree of rotational χ_1/χ_2 flexibility associated with this side chain. For example, in

[6]That is to say, the upper limit for the distance between a hydrogen bond donor (heavy) atom and the acceptor atom used by WHAT IF (*64*, *65*).

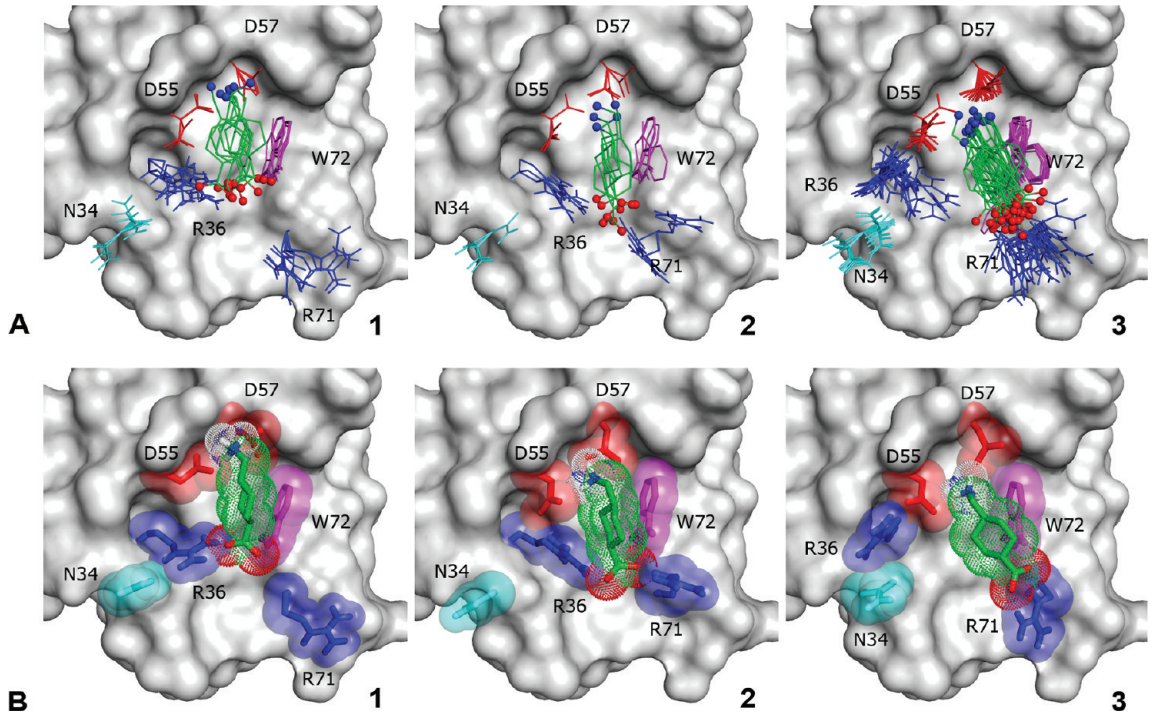


FIGURE 8: Lysine-binding site of K57D kringle 3: sampling of docked AMCHA poses. Side chains from Asp55 and Asp57 (acidic, red), Arg36 and Arg71 (basic, blue), Asn34 (cyan), and Trp72 (magenta) were allowed rotational flexibility during docking; other residues (gray surface) were kept rigid. Flexible atoms, including ligand backbone (green) and cationic (N atom, blue) and anionic (O atoms, red) polar groups, are in stick representation. (A) Superposition of individual docked conformations. Only poses within 2 Å rmsd from the lowest energy structure are shown; nonrigid hydrogen atoms are omitted. Ligand polar groups are indicated by spheres. Columns segregate the preferred K3 cationic center interacting with the ligand (see text): (A1) only Arg36 (~30.8% of dockings), (A2) Arg36 and Arg71 combined (~6.7% of dockings), and (A3) only Arg71 (~62.5% of dockings). (B) Lowest energy docked conformation corresponding to clusters A1, A2, and A3. van der Waals surfaces are superposed to all flexible atoms and distinguish between side chains (labeled) and ligand (represented by spheres and dots, respectively); polar hydrogen atoms on the ligand are shown (white).

Table 3: Binding Site Average Polar Distances (Å) between Docked AMCHA and K57D Kringle 3

distance	cluster 1	cluster 2	cluster 3
AMCHA-N···C^{γ}-Asp55	4.1 ± 1.0	4.0 ± 1.0	4.8 ± 1.8
AMCHA-N···C^{γ}-Asp57	4.8 ± 2.0	3.5 ± 0.8	4.0 ± 1.4
AMCHA-C(9)···$N^{\delta 2}$-Asn34	7.4 ± 2.1	9.3 ± 1.6	11.7 ± 2.2
AMCHA-C(9)···$N^{\eta 1}/N^{\eta 2}$-Arg36 [a]	3.5 ± 0.3	3.7 ± 0.3	8.4 ± 1.8
AMCHA-C(9)···$N^{\eta 1}/N^{\eta 2}$-Arg71	10.6 ± 2.8	3.6 ± 0.3	3.4 ± 0.2

[a]For ambiguous atom sets, the lower observed distance is listed.

hPgn K1, site 72 is conservatively substituted by a Tyr residue (Scheme 1A). In addition, when Trp72 in hPgn K4 was selectively reacted with [^{14}C]dimethyl(2-hydroxy-5-nitrobenzyl)sulfonium bromide, it eliminated the lysine-binding affinity (*101*). An animation sequentially showing independent docking results with flexible K3 side chains is included as Supporting Information, Figure S3.

Phylogenetic Analysis. Comparison of Pgn K3 sequences from a number of species[7] (Scheme 2A) reveals that residues Lys57 and His64 are unique to the human variant. Indeed, in all other species Pgn K3 consistently fills site 57 with a Glu residue, similar to hPgn K2 but in contrast to hPgn K1, K4, and K5 that carry an Asp at site 57, and an aromatic residue at site 64 [a Tyr in all species except in *Erinaceus europaeus* (ERIEU,[8] European hedgehog), where a Phe fills the site].

Curiously, the ERIEU apolipoprotein(a) [Apo(a)] (*106*) contains a sequential array of 31 nearly identical copies of Pgn K3[9] in a pearl necklace-like configuration. It is striking that residues Glu57 and Tyr64 are strictly conserved in these paralogous K3 repeats (Scheme 2B). More recently, an ERIEU apolipoprotein-related protein (Arp) was found (*105*) to contain seven K3-type domains which fill site 57 with a Ser residue (Scheme 2B); additionally, with the exception of the first and fifth kringle, all replace the heretofore strictly conserved Asp55 with a Ser residue (*105*). While less acidic than the Asp or Glu residues that occupy this position in lysine-binding kringle domains such as hPgn K1, K2, K4, and K5 (Scheme 1B), Ser is nonetheless weakly polar and hydrophilic. In contrast, the Asp57/Glu57 → Lys substitution in the hPgn K3 represents a negative-to-positive charge replacement at the LBS that does not occur either in the other hPgn kringles or in Pgn K3 sequences from other species or even in the ERIEU Apo(a) and Arp.

Another salient feature of the Pgn K3 domain is occurrence of the Cys43 residue, involved in the K2(Cys4)−(Cys43)K3 inter-kringle cystine link: it is strictly conserved in all species (Scheme 2A) and is also found in the last K3-type repeat of ERIEU Apo(a) (Scheme 2B). In the latter it has been suggested (*106*) to be responsible for covalent linkage to Apo(b)

[7]Confer Protein Information Resource (*102*) (http://pir.georgetown.edu) entry PIRSF001150.

[8]We denote organisms by the mnemonic used in the NEWT taxonomy portal (*103*).

[9]In contrast, the human Apo(a) [Uniprot (*104*) access number P08519 (*95*)] consists of multiple copies of K4-type units [individually referred to as K4(i)] with one copy of K5 preceding the C-terminus, a case of convergent evolution (*94*, *105*, *106*). By analogy, we refer to non-Pgn K3-type domains as K3(i).

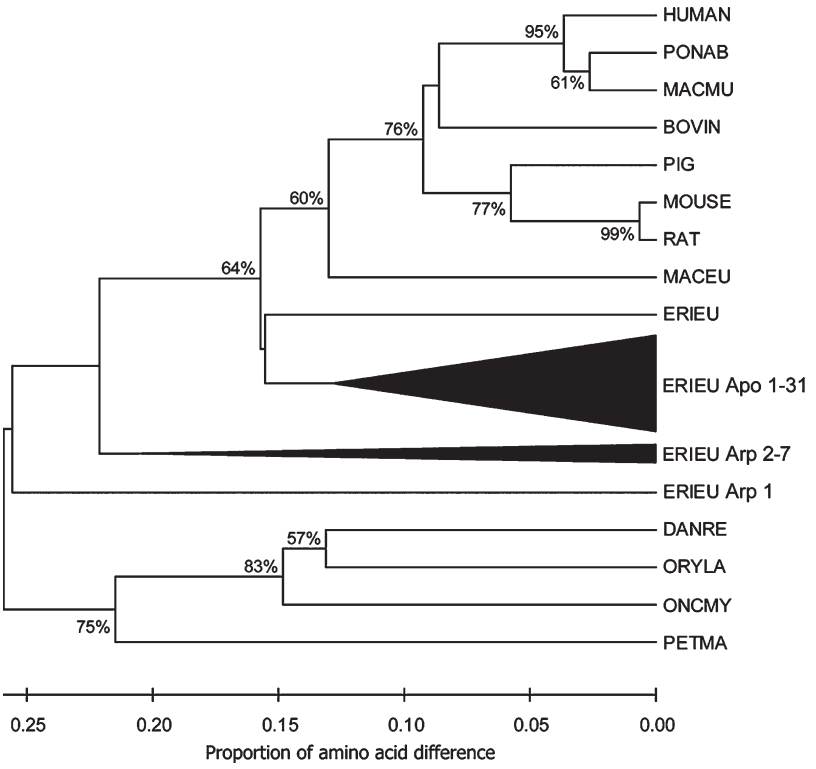


FIGURE 9: Kringle 3 phylogenetic tree for the set of plasminogen domains and paralogue repeats. Bootstrap consensus tree (*91*) of sequences in Scheme 2 (51 taxa; see text), inferred from 100000 replicates of a dendrogram generated using the unweighted pair-group method with arithmetic mean (UPGMA) (*90*). Bootstrap support ≥ 50% is indicated next to the branches. Collapsed manifolds corresponding to individual ERIEU Apo and Arp kringles are shown as black triangles; all other taxa refer to Pgn. Phylogenetic analyses were via MEGA4 (*88*). Evolutionary distances, corrected using the Poisson method (*92*), are drawn to scale based on the proportion of amino acid difference at 78 aligned residue positions.

by analogy to the unpaired Cys69 residue in the penultimate K4-type kringle of human Apo(a) (*95*, *107*). In the Arp K3(1), a single free Cys residue is found in yet another position, site 52 (Scheme 2B), thereby replacing Arg52 in what otherwise would be a strictly conserved area. However, its buried location in the core of the domain makes it unlikely that the Arg52 → Cys substitution might have arisen to mediate covalent coupling to another protein.

A phylogenetic tree constructed for 51 homologous K3-type domains is shown in Figure 9. It is reassuring that the branching pattern reflects the accepted taxonomic evolution of vertebrates, most noteworthy in that it segregates mammalian Pgn K3 sequences from those of ray-finned fish species, *Oncorhynchus mykiss* (ONCMY, rainbow trout), *Danio rerio* (DANRE, zebrafish), and *Oryzias latipes* (ORYLA, Japanese ricefish), themselves branched out with good (75%) bootstrap support from *Petromyzon marinus* (PETMA, sea lamprey (*108*)). The latter, evolutionary one of the earliest true vertebrates, represents the most "primitive" animal known to have a hemostatic system (*109*). The phylogenetic tree (Figure 9) is in line (64% bootstrap support) with earlier speculations: while the remodeling and duplication of Pgn K3 leading to the ERIEU Apo(a) may have predated placental mammal radiation (*110*), and perhaps even divergence of marsupial and placental mammals (*111*) (60% bootstrap support), it nonetheless followed the evolution of terrestrial vertebrates.

The tree suggests that Arp kringles diverged from the main K3 branch earlier than those in Apo(a) (Figure 9) and hence are less closely related to Pgn K3 (compare Scheme 2A with the consensus sequence in Scheme 2B). Interestingly, while Arp K3(1) clusters separately from the other Arp kringle repeats [Arp K3(2)−K3(7)], a test by which the free Cys52 in the former domain was replaced by the consensus Arg52 caused the phylogenetic branching to collapse. The codon usage table for ERIEU (*112*) implies that, notwithstanding two Arg codons differing only in one nucleotide from the corresponding Cys pair, the usage bias is substantially different (~7.7 ‰ vs ~20.5 ‰). Both observations suggest that the unlikely Arg52 → Cys substitution took place under considerable evolutionary pressure; alternatively, one may speculate that Cys52 in Arp K3(1) might be a leftover from a "proto-kringle" domain (*113*), before evolutionary appearance of the structurally crucial central Cys22−Cys63/Cys51−Cys75 disulfide "cross", a signature of the kringle fold (*23*, *24*).

Ligand Binding to K3 Variants. The evolutionary diversity of the K3-type domains affords possibilities for additional *in silico* docking tests in order to assess the potential role of specific amino acid residues in the K3−ligand interactions. The effective affinities (K_a^*) of the individual investigated kringle/zwitterionic ligand combinations, estimated from eq 4, are provided in Supporting Information, Table S2; Figure 10 highlights the ligand-averaged *relative* affinities ($\overline{K_a^*}$) obtained by averaging K_a^* values over all four investigated zwitterionic ligands.

(*a*) *Synthetic hPgn K3 Mutants*. Codon pairs encoding for Asp vs Glu and Glu vs Lys differ respectively in single base mutations while maintaining (in humans) a similar codon usage bias (*112*). An analogous progression is observed for the Phe, Tyr, and His codons (*112*). Thus, it is tempting to speculate that hPgn K3 (Scheme 1A) evolved from a kringle precursor in at least two distinct steps: first with the Asp57 → Glu substitution observed in the Pgn K3 from all other species (Scheme 2A) and second with a unique and concerted double substitution Glu57 → Lys/Tyr64 → His. Such an evolutionary path would corroborate

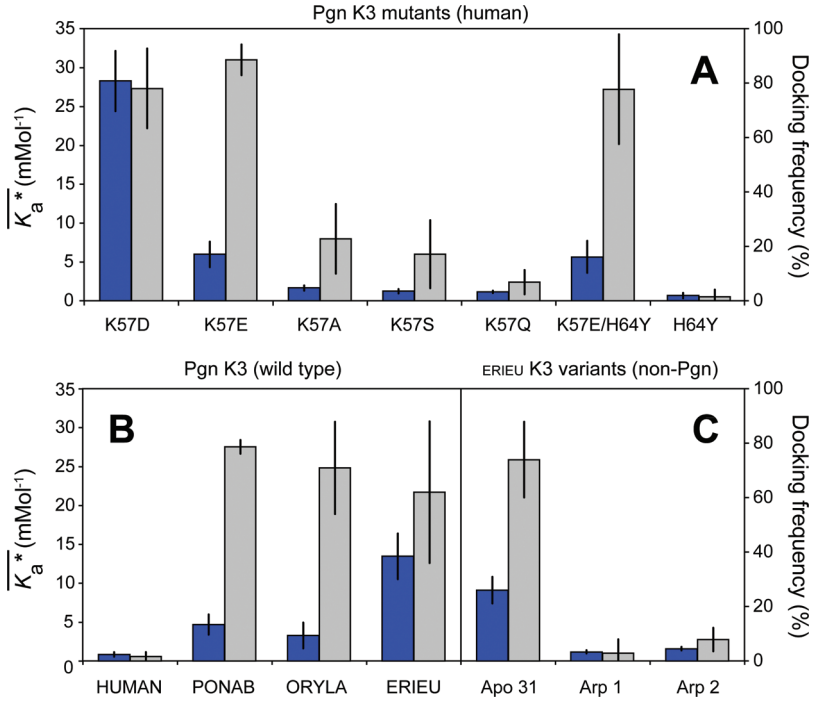


FIGURE 10: AutoDock estimated relative zwitterionic ligand affinities of kringle 3 domains. $\overline{K_a^*}$ values (blue) and docking frequencies at the canonical LBS (gray) are averages over all ligands (extracted from Supporting Information, Table S2). Data dispersion are indicated by error bars. (A) Impact of amino acid point mutations. (B) Orthologues and (C) ERIEU paralogues. Species and protein labels are as defined in Scheme 2.

the hypothesis that amino acids essential for the domain fold are generally more conserved than those related with its function (*27*, *114*); the proposed path also receives a measure of support from the genealogy of kringle-related structures (*27*, *113*).

Besides the Glu57 → Lys substitution, the genetic code suggests other, potentially less disruptive replacements for this amino acid; both the Glu57 → Ala and Glu57 → Gln substitutions would only require a single base mutation (*112*). The four ligands tested previously (6AHA, AcLys, AMCHA, and BASA) were *in silico* docked to rigid models of several mutated K3 constructs in order to further investigate the impact of the locus 57 residue on the lysine-binding capability of the domain. Inclusion of an Asp residue at site 57 results in the most potent LBS overall, comparing favorably to the affinity conferred by Glu57 (Figure 10A) usually found in nonhuman Pgn K3 domains (Scheme 2A). This would be in line with experimental findings for K2, where the double mutation R56G/E57D resulted in an overall *increased* affinity for zwitterionic ligands (*29*). It was previously suggested that the Asp side chain, which is one methylene unit shorter than Glu, would place the ligand backbone in closer contact with the aromatic side chains that line the LBS hydrophobic groove (*29*). Interestingly, our results suggest that while a negatively charged residue such as Asp or Glu is undoubtedly required for an optimally active LBS, replacement of residue 57 for Ala, Ser, or Gln maintains a degree of affinity for zwitterionic ligands. Notably, for all of these site 57 variants, the preference of the preformed binding site in K3 for cyclic over linear zwitterionic ligands is preserved. As indicated by the solution structure of rK3, Lys57 both disrupts the binding site charge distribution and partially obstructs access to the hydrophobic groove, thus sterically preventing zwitterionic ligand binding. Indeed, Arp, which contains a Ser at position 57 in each of its seven kringle domains, is retained by Lys-conjugated chromatographic resins (*105*), whereas the wild-type hPgn K3 is not (*30*).

While the Glu57 → Lys substitution is of significant functional consequences, H64Y-mutated hPgn K3 shows only a slight propensity to interact with AcLys and, marginally, 6-AHA (Supporting Information, Table S2) with very few dockings at the binding site proper (4.1% and 0.8%, respectively). Furthermore, the double mutation K57E/H64Y (Figure 11B), which replicates the configuration of K3 in other species, results in only a slight variation of the overall affinity for zwitterionic ligands compared to the single K57E mutation (Figure 10A) with the calculated $\overline{K_a^*}$ values remaining unchanged within the margin of error (Supporting Information, Table S2).

(*b*) *K3 Orthologues*. The above results suggest that variability among the orthologous Pgn K3 sequences (Scheme 2A) is unlikely to significantly alter the protein fold. Furthermore, the analysis also hints at a degree of structural/functional adaptability of the canonical K3 binding site toward residue substitutions. Indeed, while the functional requirement of site 57 for negatively charged residues has long been recognized (*18*, *27*, *29*, *99*, *115*), the sequence of Arp kringles points to a heretofore unnoticed tolerance of this locus for Ser, a residue that carries an uncharged, albeit hydrophilic, side chain. We therefore explored the potential affinity of K3 species variants to a set of zwitterionic ligands. The obtained homology-based models for *Pongo abelii* (PONAB, Sumatran orangutan), ORYLA, and ERIEU suggest in all cases a definite, albeit weak, affinity for the tested zwitterionic ligands (Figure 10B and Supporting Information, Table S2). This result, while not unexpected as these domains all contain Glu57 (Scheme 2A), highlights the structural exceptionality of the human Pgn K3 domain, whose binding site lacks the canonical anionic center (Figure 4, Figure 11A) found in LBS's of all other Lys-binding kringles (Figure 11, panels B−D).

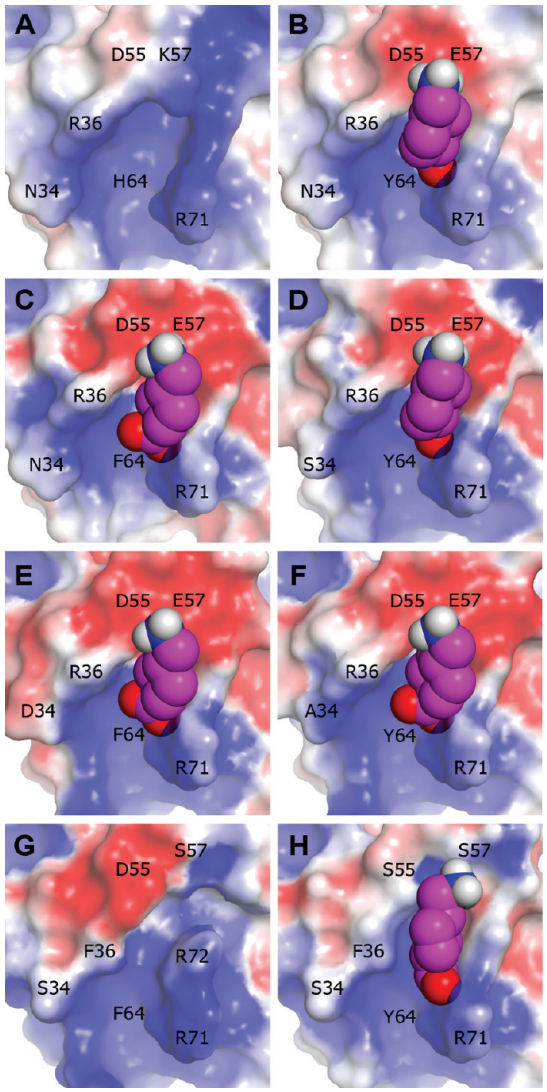


FIGURE 11: AMCHA docking to kringle 3 type domains. Structures are generated by homology modeling based on the NMR structure of hPgn K3. Kringle electrostatic surfaces are calculated and colored as in Figure 6C; selected residues are labeled. Binding sites with suggested affinity to zwitterionic ligands are shown with docked AMCHA in effective van der Waals radii, space-filling representation. The pose of the ligand was predicted by *in silico* docking via AutoDock 4.0 on the rigid kringle: (A) wild-type HUMAN Pgn K3; (B) K57E/H64Y HUMAN Pgn K3; (C) PONAB Pgn K3; (D) ORYLA Pgn K3; (E) ERIEU Pgn K3; (F) ERIEU Apo K3(31); (G) ERIEU Arp K3(1); (H) ERIEU Arp K3(2).

Interestingly, our results suggest that the ERIEU Pgn K3 may be endowed with a higher affinity for Lys-like zwitterions than either the PONAB or the ORYLA orthologues (Figure 10B). Comparison of the amino acid sequence (Scheme 2A) reveals two unusual substitutions in ERIEU, namely, Tyr64 → Phe and Asn34 → Asp: while the former represents a conservative substitution, the latter is significant as it contributes a negative electrostatic component at the lower fringe of the binding site (Figure 11E). Thus, the Asn34 → Asp substitution would be expected to drive the ligand anionic group to place itself closer to Arg71 at the canonical LBS cationic center. Notably, docking tests on a model of human (K57D)K3 derived from the X-ray structure, which includes a similar N34E mutation, predict an increased affinity for zwitterionic ligands (Table 2). Furthermore, in hPgn K3, Asn34 undergoes a posttranslational N-linked glycosylation (*116*), the effects of which may affect binding capability and circulatory half-time (*117*). In contrast, the ERIEU Pgn K3 lacks a glycosylation site (*118*). In this context, the postulated "recurring evolution" of Apo(a) (*106*, *111*, *119*) is most telling: while the ERIEU Pgn K3 might have filled the role of a precursor for the ERIEU fibrin-binding Apo(a) (*106*), in primates the K3 is unlikely to have served such a purpose (*106*). Indeed, the corresponding Apo(a) in both humans and *Macaca mulatta* (macmu, Rhesus macaque) are thought to have evolved from the Pgn K4 and K5 modules instead (*113*). We speculate that the ERIEU Pgn K3's unique position as a likely precursor for Apo(a) may be facilitated by the singular presence of Asp34. In particular, it is suggested that Asp34 has a dual effect in that it appears to enhance zwitterionic ligand affinity (Figure 10B) while negating the potential for N-linked glycan formation, in contrast to all other mammalian Pgn K3 sequences (Scheme 2A).

(*c*) *K3 Paralogues*. The ERIEU Apo(a) K3(31) is the Apo(a) kringle sequence most closely related to that of the ERIEU Pgn K3 (*106*, *110*); significantly, the two are suggested to display similar affinities toward zwitterionic ligands, both in terms of binding energies (Supporting Information, Table S2) and in LBS ligand pose as predicted by the docking of AMCHA (Figure 11E,F). This is in line with the hypothesis, borne from phylogenetic analyses, that ERIEU Apo(a) evolved independently of primate Apo(a) by repeated duplication of the gene encoding for Pgn K3 (*110*, *111*, *119*, *120*).

Human Apo(a) has been proposed to interfere with activation of plasminogen (*95*), thereby contributing to reduced fibrin clot lysis. Independently, it has been shown that ERIEU Apo(a) binds to Lys and to immobilized fibrin surfaces (*121*). Therefore, both human and ERIEU Apo(a) are likely to regulate lysine-mediated proteolysis in what appears to be, at least in case of ERIEU, a seasonal physiological phenomenon (*121*).

The first 21 kringle domains in ERIEU Apo(a) can be clustered into two sets based on their ordinal parity (even/odd numbering), a pattern best explained by multiple duplication of a tandem kringle unit in the course of evolution (*110*). Inspection of the amino acid sequence (Scheme 2B) shows that residues in the canonical LBS, in particular Arg36 and Asp55, are more conserved in odd-numbered kringles than in their even-numbered counterparts and suggests that the affinity for zwitterions may be more pronounced in the former. Simultaneously, the even-numbered kringles of Apo(a) bear similarities with the kringles in Arp, including replacement of Arg36 by an aromatic residue (Scheme 2B). Combined with the Glu57 → Ser replacement discussed above, the even-numbered kringles are expected to show weakened binding capability. Our docking simulations with homology models of Arp K3(1) and Arp K3(2) kringles support this hypothesis (Figure 11G,H). Indeed, Arp K3(2) is suggested to exhibit a weak *in silico* affinity for zwitterionic ligands (Figure 10C) comparable to that predicted for hPgn (K57S)K3 mutant (Figure 10A). On a similar basis, the Arp K3(1) would behave *in silico* similarly to hPgn (H64Y)K3 in that it is predicted to exhibit a weak affinity for AcLys not replicated by the other ligands. Individual Arp kringles are hence unlikely to contribute to Lys binding. Thus, our results concur with the suggestion that the reported capability of the Arp protein to interact with Lys-Sepharose (*105*) may result from cooperative binding that involves several of these constructs.

Although Arp K3(1) retains in Asp55 a canonical LBS acidic residue (Scheme 2B), it would seem to exhibit almost *nil* affinity for zwitterionic ligands (Figure 10C and Supporting Information, Table S3). This is not surprising as in Arp K3(1), the aromatic

residue at site 72 is replaced by Arg (Scheme 2B, Figure 11G). Indeed, while conferring extra basicity to the binding site, Arg72 is suggested to sterically interfere with *in silico* docking of prototype zwitterionic ligands by affecting a critical hydrophobic interaction (*101*).

It has been noted that "gene duplication leading to functional novelty is ordinarily preceded by bifunctionality of the ancestral gene" (*110*). Hence K3, and with regard to primate Apo(a), K4 and K5, may concur with the even-numbered kringles of ERIEU Apo(a) and the "aberrant" kringles of Arp as affording examples of binding sites that have evolved from a bifunctional kringle protodomain toward optimization of a function that differs from that of canonical Lys-mediated fibrin binding. Indeed, recent studies in our laboratories have shown that hPgn K3, K5, and, to a lesser extent, K4 exhibit affinity for fatty acids (*16*, *122*). This would be in line with the repetitive role of these kringles as building blocks in Apo(a)-type constructs, which in plasma wrap around lipid aggregates (*95*). Interestingly, preliminary docking tests (unpublished) predict that oleic acid binds to ERIEU Apo K3(31), Arp K3(2) and Arp K3(1).

CONCLUSIONS

The NMR analysis of hPgn K3 reported here provides both the solution structure of the domain and, *a fortiori*, the assignment of the spectroscopic connectivities that become specifically mapped onto the fold. Indeed, structure and spectra assignment are equivalent: knowledge of the one implies the other (*123*, *124*). Therefore, ligand-induced spectroscopic perturbation leads to the localization, at atomic level resolution, of the ligand-binding site. Such information proves to be fundamental for the functional characterization of both the wild type and the (K57D)K3 mutant and justifies solving the NMR solution structure as a mandatory prerequisite for the functional characterization *even if a crystallographic structure is available*.

Via a combined NMR/docking simulation protocol, we were able to exploit homology modeling as a basis to investigate *in silico* Pgn K3 domains ranging from the Sumatran orangutan to the Japanese ricefish and that exemplify the highest and the lowest known homology, respectively, with the human K3 sequence. Thus, while coarse-grained, a picture of the structural/functional evolutionary connectivities among selected Pgn K3 variants could be achieved. The European hedgehog, in particular, provides a rich source of K3 variants that enabled us to tentatively test the role of specific binding site residues in the potential complexation of zwitterionic ω-amino acid ligands, thus shedding light on the hypothetical physiological chemistry of the ERIEU Apo(a) in-tandem multikringle construct.

It is our hope that this study shall serve to foster the combined use of NMR and *blind, unrestricted* docking experiments in the characterization of potential protein−effector interactions, whether in their biological milieu or in other pharmacological contexts.

ACKNOWLEDGMENT

NMR spectra were recorded and processed by S. Gunawardena. The authors thank U. Kämpfer (University of Bern) for valuable help in preparing and characterizing the protein samples and V. Simplaceanu (Carnegie Mellon) for expert NMR advice and assistance.

SUPPORTING INFORMATION AVAILABLE

A table summarizing the implemented experimental NMR constraints and the extent of agreement by the reported structural bundle, a table listing individual effective association constants and docking frequencies at the canonical binding site for the interaction of zwitterionic ligands to K3 variants, and three additional figures as described in the text: a graphic summary of K3 NMR sequential connectivities and dihedral constraints; results for docking zwitterionic ligands to wild-type K3; an animation of individual snapshots for docking AMCHA to (K57D)K3 with flexible Asn34, Arg36, Asp55, Asp57, Arg71, and Trp72 side chains. This material is available free of charge via the Internet at http://pubs.acs.org.

REFERENCES

1. Schaller, J., Gerber, S., Kämpfer, U., Lejon, S., and Trachsel, C. (2008) Human Blood Plasma Proteins, Structure and Function, John Wiley & Sons, Chichester, England.
2. Castellino, F. J., and Ploplis, V. A. (2005) Structure and Function of the Plasminogen/Plasmin System. *Thromb. Haemostasis 93*, 647–654.
3. Mignatti, P., and Rifkin, D. B. (1993) Biology and Biochemistry of Proteinases in Tumor Invasion. *Physiol. Rev. 73*, 161–195.
4. Sottrup-Jensen, L., Claeys, H., Zajdel, M., Petersen, T. E., and Magnusson, S. (1978) The Primary Structure of Human Plasminogen: Isolation of Two Lysine-Binding Fragments and One "Mini-" Plasminogen (Mw 38,000) by Elastase-Catalyzed-Specific Limited Proteolysis. *Prog. Chem. Fibrinolysis Thrombolysis 3*, 191–209.
5. Tordai, H., Bányai, L., and Patthy, L. (1999) The PAN Module: The N-Terminal Domains of Plasminogen and Hepatocyte Growth Factor Are Homologous with the Apple Domains of the Prekallikrein Family and with a Novel Domain Found in Numerous Nematode Proteins. *FEBS Lett. 461*, 63–67.
6. Douglas, J. T., von Haller, P. D., Gehrmann, M. L., Llinás, M., and Schaller, J. (2002) The Two-Domain NK1 Fragment of Plasminogen: Folding, Ligand Binding, and Thermal Stability Profile. *Biochemistry 41*, 3302–3310.
7. Magnusson, S., Sottrup-Jensen, L., Petersen, T. E., Dudek-Wojciechowska, G., and Claeys, H. (1976) Homologous "Kringle" Structures Common to Plasminogen and Prothrombin. Substrate Specificity of Enzymes Activating Prothrombin and Plasminogen. *Miami Winter Symp. 11*, 203–238.
8. Mathews, I. I., Vanderhoff-Hanaver, P., Castellino, F. J., and Tulinsky, A. (1996) Crystal Structures of the Recombinant Kringle 1 Domain of Human Plasminogen in Complexes with the Ligands ε-Aminocaproic Acid and Trans-4-(Aminomethyl)Cyclohexane-1-Carboxylic Acid. *Biochemistry 35*, 2567–2576.
9. Rejante, M. R., and Llinás, M. (1994) Solution Structure of the ε-Aminohexanoic Acid Complex of Human Plasminogen Kringle 1. *Eur. J. Biochem. 221*, 939–949.
10. Wu, T. P., Padmanabhan, K. P., and Tulinsky, A. (1994) The Structure of Recombinant Plasminogen Kringle 1 and the Fibrin Binding Site. *Blood Coagulation Fibrinolysis 5*, 157–166.
11. Marti, D. N., Schaller, J., and Llinás, M. (1999) Solution Structure and Dynamics of the Plasminogen Kringle 2-AMCHA Complex: 3(1)-Helix in Homologous Domains. *Biochemistry 38*, 15741–15755.
12. Rios-Steiner, J. L., Schenone, M., Mochalkin, I., Tulinsky, A., and Castellino, F. J. (2001) Structure and Binding Determinants of the Recombinant Kringle-2 Domain of Human Plasminogen to an Internal Peptide from a Group a Streptococcal Surface Protein. *J. Mol. Biol. 308*, 705–719.
13. Atkinson, R. A., and Williams, R. J. (1990) Solution Structure of the Kringle 4 Domain from Human Plasminogen by 1H Nuclear Magnetic Resonance Spectroscopy and Distance Geometry. *J. Mol. Biol. 212*, 541–552.
14. Wu, T. P., Padmanabhan, K., Tulinsky, A., and Mulichak, A. M. (1991) The Refined Structure of the ε-Aminocaproic Acid Complex of Human Plasminogen Kringle 4. *Biochemistry 30*, 10589–10594.
15. Chang, Y., Mochalkin, I., McCance, S., Cheng, B., Tulinsky, A., and Castellino, F. J. (1998) Structure and Ligand Binding Determinants of the Recombinant Kringle 5 Domain of Human Plasminogen. *Biochemistry 37*, 3258–3271.
16. Battistel, M. D., Grishaev, A., An, S. S. A., Castellino, F. J., and Llinás, M. (2009) Solution Structure and Functional Characteriza-

tion of Human Plasminogen Kringle 5. *Biochemistry 48*, 10208–10219.
17. Padmanabhan, K., Wu, T. P., Ravichandran, K. G., and Tulinsky, A. (1994) Kringle-Kringle Interactions in Multimer Kringle Structures. *Protein Sci. 3*, 898–910.
18. Byeon, I. J., and Llinás, M. (1991) Solution Structure of the Tissue-Type Plasminogen Activator Kringle 2 Domain Complexed to 6-Aminohexanoic Acid, an Antifibrinolytic Drug. *J. Mol. Biol. 222*, 1035–1051.
19. Gehrmann, M. L. (2002) Structural and Functional Similarities between FII and Kringle Domains, *Doctoral Dissertation*, Carnegie Mellon University, Pittsburgh, PA.
20. Li, X., Smith, R., and Dobson, C. (1992) Sequential Proton NMR Assignments and Secondary Structure of the Kringle Domain from Urokinase. *Biochemistry 31*, 9562–9571.
21. Barinka, C., Parry, G., Callahan, J., Shaw, D., Kuo, A., Bdeir, K., Cines, D., Mazar, A., and Lubkowski, J. (2006) Structural Basis of Interaction between Urokinase-Type Plasminogen Activator and Its Receptor. *J. Mol. Biol. 363*, 482–495.
22. Abad, M. C., Arni, R. K., Grella, D. K., Castellino, F. J., Tulinsky, A., and Geiger, J. H. (2002) The X-Ray Crystallographic Structure of the Angiogenesis Inhibitor Angiostatin. *J. Mol. Biol. 318*, 1009–1017.
23. Tulinsky, A., Park, C. H., Mao, B., and Llinás, M. (1988) Lysine/Fibrin Binding Sites of Kringles Modeled after the Structure of Kringle 1 of Prothrombin. *Proteins 3*, 85–96.
24. Ozhogina, O. A., and Bominaar, E. L. (2009) Characterization of the Kringle Fold and Identification of a Ubiquitous New Class of Disulfide Rotamers. *J. Struct. Biol. 168*, 223–233.
25. Kapetanopoulos, A., Fresser, F., Millonig, G., Shaul, Y., Baier, G., and Utermann, G. (2002) Direct Interaction of the Extracellular Matrix Protein Dance with Apolipoprotein(a) Mediated by the Kringle IV-Type 2 Domain. *Mol. Genet. Genomics 267*, 440–446.
26. Lee, B. H., Kim, H.-K., and Joe, Y. A. (2008) The Kringle Domain of Tissue-Type Plasminogen Activator Inhibits Extracellular Matrix-Induced Adhesion and Migration of Endothelial Cells. *Biosci., Biotechnol., Biochem. 72*, 2303–2308.
27. Patthy, L., Trexler, M., Váli, Z., Bányai, L., and Varadi, A. (1984) Kringles: Modules Specialized for Protein Binding. Homology of the Gelatin-Binding Region of Fibronectin with the Kringle Structures of Proteases. *FEBS Lett. 171*, 131–136.
28. Mahdy, A. M., and Webster, N. R. (2004) Perioperative Systemic Haemostatic Agents. *Br. J. Anaesth. 93*, 842–858.
29. Marti, D. N., Hu, C. K., An, S. S., von Haller, P., Schaller, J., and Llinás, M. (1997) Ligand Preferences of Kringle 2 and Homologous Domains of Human Plasminogen: Canvassing Weak, Intermediate, and High-Affinity Binding Sites by ^{1}H-NMR. *Biochemistry 36*, 11591–11604.
30. Marti, D. N., Schaller, J., Ochensberger, B., and Rickli, E. E. (1994) Expression, Purification and Characterization of the Recombinant Kringle 2 and Kringle 3 Domains of Human Plasminogen and Analysis of Their Binding Affinity for ω-Aminocarboxylic Acids. *Eur. J. Biochem. 219*, 455–462.
31. Söhndel, S., Hu, C. K., Marti, D. N., Affolter, M., Schaller, J., Llinás, M., and Rickli, E. E. (1996) Recombinant Gene Expression and ^{1}H NMR Characteristics of the Kringle (2 + 3) Supermodule: Spectroscopic/Functional Individuality of Plasminogen Kringle Domains. *Biochemistry 35*, 2357–2364.
32. Bürgin, J., and Schaller, J. (1999) Expression, Isolation and Characterization of a Mutated Human Plasminogen Kringle 3 with a Functional Lysine Binding Site. *Cell. Mol. Life Sci. 55*, 135–141.
33. Forsgren, M., Raden, B., Israelsson, M., Larsson, K., and Hedén, L. O. (1987) Molecular Cloning and Characterization of a Full-Length cDNA Clone for Human Plasminogen. *FEBS Lett. 213*, 254–260.
34. Zamenhof, P. J., and Villarej, M. (1972) Construction and Properties of *Escherichia coli* Strains Exhibiting α-Complementation of β-Galactosidase Fragments in-Vivo. *J. Bacteriol. 110*, 171–178.
35. Frank, P. S., Douglas, J. T., Locher, M., Llinás, M., and Schaller, J. (2003) Structural/Functional Characterization of the α2-Plasmin Inhibitor C-Terminal Peptide. *Biochemistry 42*, 1078–1085.
36. Aue, W. P., Bartholdi, E., and Ernst, R. R. (1976) Two-Dimensional Spectroscopy. Application to Nuclear Magnetic Resonance. *J. Chem. Phys. 64*, 2229–2246.
37. Braunschweiler, L., and Ernst, R. R. (1983) Coherence Transfer by Isotropic Mixing—Application to Proton Correlation Spectroscopy. *J. Magn. Reson. 53*, 521–528.
38. Jeener, J., Meier, B. H., Bachmann, P., and Ernst, R. R. (1979) Investigation of Exchange Processes by 2-Dimensional NMR Spectroscopy. *J. Chem. Phys. 71*, 4546–4553.
39. Bodenhausen, G., and Ruben, D. J. (1980) Natural Abundance Nitrogen-15 NMR by Enhanced Heteronuclear Spectroscopy. *Chem. Phys. Lett. 69*, 185–189.
40. Talluri, S., and Wagner, G. (1996) An Optimized 3D NOESY-HSQC. *J. Magn. Reson. B 112*, 200–205.
41. Düx, P., Whitehead, B., Boelens, R., Kaptein, R., and Vuister, G. W. (1997) Measurement of ^{15}N-^{1}H Coupling Constants in Uniformly ^{15}N-Labeled Proteins: Application to the Photoactive Yellow Protein. *J. Biomol. NMR 10*, 301–306.
42. Archer, S. J., Ikura, M., Torchia, D. A., and Bax, A. (1991) An Alternative 3D NMR Technique for Correlating Backbone ^{15}N with Side Chain Hβ Resonances in Larger Proteins. *J. Magn. Reson. 95*, 636–641.
43. Piotto, M., Saudek, V., and Sklenár, V. (1992) Gradient-Tailored Excitation for Single-Quantum NMR Spectroscopy of Aqueous Solutions. *J. Biomol. NMR 2*, 661–665.
44. Bax, A., and Davis, D. G. (1985) Mlev-17-Based Two-Dimensional Homonuclear Magnetization Transfer Spectroscopy. *J. Magn. Reson. 65*, 355–360.
45. Rucker, S. P., and Shaka, A. J. (1989) Broad-Band Homonuclear Cross Polarization in 2D NMR Using DIPSI-2. *Mol. Phys. 68*, 509–517.
46. De Marco, A. (1977) pH Dependence of Internal References. *J. Magn. Reson. 26*, 527–528.
47. Wüthrich, K. (1990) Protein Structure Determination in Solution by NMR Spectroscopy. *J. Biol. Chem. 265*, 22059–22062.
48. Marion, D., Driscoll, P. C., Kay, L. E., Wingfield, P. T., Bax, A., Gronenborn, A. M., and Clore, G. M. (1989) Overcoming the Overlap Problem in the Assignment of Proton NMR Spectra of Larger Proteins by Use of Three-Dimensional Heteronuclear Proton-Nitrogen-15 Hartmann-Hahn-Multiple Quantum Coherence and Nuclear Overhauser-Multiple Quantum Coherence Spectroscopy: Application to Interleukin 1β. *Biochemistry 28*, 6150–6156.
49. Marion, D., Kay, L. E., Sparks, S. W., Torchia, D. A., and Bax, A. (1989) Three-Dimensional Heteronuclear NMR of Nitrogen-15 Labeled Proteins. *J. Am. Chem. Soc. 111*, 1515–1517.
50. Llinás, M., De Marco, A., Hochschwender, S. M., and Laursen, R. A. (1983) A Proton NMR Study of Isolated Domains from Human Plasminogen. Structural Homology between Kringles 1 and 4. *Eur. J. Biochem. 135*, 379–391.
51. Vranken, W. F., Boucher, W., Stevens, T. J., Fogh, R. H., Pajon, A., Llinás, M., Ulrich, E. L., Markley, J. L., Ionides, J., and Laue, E. D. (2005) The CCPN Data Model for NMR Spectroscopy: Development of a Software Pipeline. *Proteins: Struct., Funct., Bioinf. 59*, 687–696.
52. Wüthrich, K. (1986) NMR of Proteins and Nucleic Acids, John Wiley & Sons, Ltd., New York.
53. Vuister, G. W., and Bax, A. (1993) Quantitative J Correlation: A New Approach for Measuring Homonuclear Three-Bond J (HNHα) Coupling Constants in ^{15}N-Enriched Proteins. *J. Am. Chem. Soc. 115*, 7772–7777.
54. Karplus, M. (1963) Vicinal Proton Coupling in Nuclear Magnetic Resonance. *J. Am. Chem. Soc. 85*, 2870–2871.
55. Hu, J.-S., and Bax, A. (1997) Determination of φ and χ1 Angles in Proteins from ^{13}C-^{13}C Three-Bond J Couplings Measured by Three-Dimensional Heteronuclear NMR. How Planar Is the Peptide Bond? *J. Am. Chem. Soc. 119*, 6360–6368.
56. Padrta, P., and Sklenar, V. (2002) Program MULDER—a Tool for Extracting Torsion Angles from NMR Data. *J. Biomol. NMR 24*, 339–349.
57. Hyberts, S. G., Märki, W., and Wagner, G. (1987) Stereospecific Assignments of Side-Chain Protons and Characterization of Torsion Angles in Eglin C. *Eur. J. Biochem. 164*, 625–635.
58. Peitsch, M. C. (1995) Protein Modeling by E-Mail. *Bio/Technology 13*, 658–660.
59. Guex, N., and Peitsch, M. C. (1997) SWISS-MODEL and the Swiss-PdbViewer. An Environment for Comparative Protein Modeling. *Electrophoresis 18*, 2714–2723.
60. Schwede, T., Kopp, J., Guex, N., and Peitsch, M. C. (2003) SWISS-MODEL: An Automated Protein Homology-Modeling Server. *Nucleic Acids Res. 31*, 3381–3385.
61. Rieping, W., Habeck, M., Bardiaux, B., Bernard, A., Malliavin, T. E., and Nilges, M. (2007) ARIA2: Automated NOE Assignment and Data Integration in NMR Structure Calculation. *Bioinformatics 23*, 381–382.
62. Brünger, A. T., Adams, P. D., Clore, G. M., DeLano, W. L., Gros, P., Grosse-Kunstleve, R. W., Jiang, J. S., Kuszewski, J., Nilges, M., and Pannu, N. S. (1998) Crystallography & NMR System: A New Software Suite for Macromolecular Structure Determination. *Acta Crystallogr., Sect. D: Biol. Crystallogr. 54*, 905–921.

63. Habeck, M., Rieping, W., Linge, J. P., and Nilges, M. (2004) NOE Assignment with ARIA 2.0: The Nuts and Bolts. *Methods Mol. Biol. 278*, 379–402.
64. Vriend, G. (1990) What If—a Molecular Modeling and Drug Design Program. *J. Mol. Graphics 8*, 52–56.
65. Hooft, R. W. W., Sander, C., and Vriend, G. (1996) Positioning Hydrogen Atoms by Optimizing Hydrogen-Bond Networks in Protein Structures. *Proteins: Struct., Funct., Genet. 26*, 363–376.
66. Richardson, J. S. (2008) *The Anatomy and Taxonomy of Protein Structure*.
67. Lovell, S. C., Word, J. M., Richardson, J. S., and Richardson, D. C. (2000) The Penultimate Rotamer Library. *Proteins: Struct., Funct., Genet. 40*, 389–408.
68. Laskowski, R. A., Rullmann, J. A. C., MacArthur, M. W., Kaptein, R., and Thornton, J. M. (1996) Aqua and Procheck-NMR: Programs for Checking the Quality of Protein Structures Solved by NMR. *J. Biomol. NMR 8*, 477–486.
69. Doreleijers, J. F., Raves, M. L., Rullmann, T., and Kaptein, R. (1999) Completeness of NOEs in Protein Structures: A Statistical Analysis of NMR Data. *J. Biomol. NMR 14*, 123–132.
70. Doreleijers, J. F., Rullmann, J. A. C., and Kaptein, R. (1998) Quality Assessment of NMR Structures: A Statistical Survey. *J. Mol. Biol. 281*, 149–164.
71. Davis, I. W., Leaver-Fay, A., Chen, V. B., Block, J. N., Kapral, G. J., Wang, X., Murray, L. W., Arendall, W. B., III, Snoeyink, J., Richardson, J. S., and Richardson, D. C. (2007) Molprobity: All-Atom Contacts and Structure Validation for Proteins and Nucleic Acids. *Nucleic Acids Res. 35*, W375–383.
72. Koradi, R., Billeter, M., and Wüthrich, K. (1996) MOLMOL: A Program for Display and Analysis of Macromolecular Structures. *J. Mol. Graphics 14*, 51–55.
73. Johnson, W. C. (1999) Analyzing Protein Circular Dichroism Spectra for Accurate Secondary Structures. *Proteins: Struct., Funct., Genet. 35*, 307–312.
74. Heinig, M., and Frishman, D. (2004) STRIDE: A Web Server for Secondary Structure Assignment from Known Atomic Coordinates of Proteins. *Nucleic Acids Res. 32*, W500–W502.
75. Baker, N. A., Sept, D., Joseph, S., Holst, M. J., and McCammon, J. A. (2001) Electrostatics of Nanosystems: Application to Microtubules and the Ribosome. *Proc. Natl. Acad. Sci. U.S.A. 98*, 10037–10041.
76. DeLano, W. L. (2002) The PyMOL Molecular Graphics System, DeLano Scientific, Palo Alto, CA.
77. Bordoli, L., Kiefer, F., Arnold, K., Benkert, P., Battey, J., and Schwede, T. (2009) Protein Structure Homology Modeling Using SWISS-MODEL Workspace. *Nat. Protoc. 4*, 1–13.
78. De Marco, A., Hochschwender, S. M., Laursen, R. A., and Llinás, M. (1982) Human Plasminogen. Proton NMR Studies on Kringle 1. *J. Biol. Chem. 257*, 12716–12721.
79. Taylor, J. R. (1997) An Introduction to Error Analysis: The Study of Uncertainties in Physical Measurements, 2nd ed., University Science Books, Sausalito, CA.
80. Goodsell, D. S., and Olson, A. J. (1990) Automated Docking of Substrates to Proteins by Simulated Annealing. *Proteins: Struct., Funct., Genet. 8*, 195–202.
81. Morris, G. M., Goodsell, D. S., Huey, R., and Olson, A. J. (1996) Distributed Automated Docking of Flexible Ligands to Proteins: Parallel Applications of AutoDock 2.4. *J. Comput.-Aided Mol. Des. 10*, 293–304.
82. Morris, G. M., Goodsell, D. S., Halliday, R. S., Huey, R., Hart, W. E., Belew, R. K., and Olson, A. J. (1998) Automated Docking Using a Lamarckian Genetic Algorithm and an Empirical Binding Free Energy Function. *J. Comput. Chem. 19*, 1639–1662.
83. Sitkoff, D., Sharp, K. A., and Honig, B. (2002) Accurate Calculation of Hydration Free Energies Using Macroscopic Solvent Models. *J. Phys. Chem. 98*, 1978–1988.
84. Dolinsky, T. J., Nielsen, J. E., McCammon, J. A., and Baker, N. A. (2004) PDB2PQR: An Automated Pipeline for the Setup of Poisson-Boltzmann Electrostatics Calculations. *Nucleic Acids Res. 32*, W665–W667.
85. Bas, D. C., Rogers, D. M., and Jensen, J. H. (2008) Very Fast Prediction and Rationalization of pKa Values for Protein-Ligand Complexes. *Proteins: Struct., Funct., Bioinf. 73*, 765–783.
86. Sanner, M. F. (1999) Python: A Programming Language for Software Integration and Development. *J. Mol. Graphics Modell. 17*, 57–61.
87. Huey, R., and Morris, G. M. (2008) Using AutoDock 4 with AutoDockTools: A Tutorial, The Scripps Research Institute, La Jolla, CA 92037−1000.
88. Tamura, K., Dudley, J., Nei, M., and Kumar, S. (2007) MEGA4: Molecular Evolutionary Genetics Analysis (MEGA) Software Version 4.0. *Mol. Biol. Evol. 24*, 1596.
89. Larkin, M., Blackshields, G., Brown, N., Chenna, R., McGettigan, P., McWilliam, H., Valentin, F., Wallace, I., Wilm, A., and Lopez, R. (2007) Clustal W and Clustal X Version 2.0. *Bioinformatics 23*, 2947.
90. Sneath, P. H. A., and Sokal, R. R. (1973) Numerical Taxonomy; the Principles and Practice of Numerical Classification, W. H. Freeman, San Francisco.
91. Felsenstein, J. (1985) Confidence Intervals on Phylogenies: An Approach Using the Bootstrap. *Evolution 39*, 783–791.
92. Zuckerkandl, E., and Pauling, L. (1965) Evolutionary Divergence and Convergence in Proteins, in Evolving Genes and Proteins (Bryson, V., and Vogel, H. J., Eds.) pp 97−166, Academic Press, New York.
93. Adzhubei, A., and Sternberg, M. (1993) Left-Handed Polyproline II Helices Commonly Occur in Globular Proteins. *J. Mol. Biol. 229*, 472–493.
94. Ye, Q., Rahman, M. N., Koschinsky, M. L., and Jia, Z. (2001) High-Resolution Crystal Structure of Apolipoprotein(a) Kringle IV Type 7: Insights into Ligand Binding. *Protein Sci. 10*, 1124–1129.
95. Koschinsky, M. L., and Marcovina, S. M. (2004) Structure-Function Relationships in Apolipoprotein(a): Insights into Lipoprotein(a) Assembly and Pathogenicity. *Curr. Opin. Lipidol. 15*, 167–174.
96. Maderegger, B., Bermel, W., Hrzenjak, A., Kostner, G. M., and Sterk, H. (2002) Solution Structure of Human Apolipoprotein(a) Kringle IV Type 6. *Biochemistry 41*, 660–668.
97. Chitayat, S., Kanelis, V., Koschinsky, M. L., and Smith, S. P. (2007) Nuclear Magnetic Resonance (NMR) Solution Structure, Dynamics, and Binding Properties of the Kringle IV Type 8 Module of Apolipoprotein(a). *Biochemistry 46*, 1732–1742.
98. Scatchard, G. (1949) The Attractions of Proteins for Small Molecules and Ions. *Ann. N.Y. Acad. Sci. 51*, 660–672.
99. Trexler, M., Váli, Z., and Patthy, L. (1982) Structure of the ω-Aminocarboxylic Acid-Binding Sites of Human Plasminogen. Arginine 70 and Aspartic Acid 56 Are Essential for Binding of Ligand by Kringle 4. *J. Biol. Chem. 257*, 7401–7406.
100. Nielsen, P. R., Einer-Jensen, K., Holtet, T. L., Andersen, B. D., Poulsen, F. M., and Thogersen, H. C. (1993) Protein-Ligand Interactions in the Lysine-Binding Site of Plasminogen Kringle 4 Are Different in Crystal and Solution. Electrostatic Interactions Studied by Site-Directed Mutagenesis Exclude Lys35 as an Important Acceptor in Solution. *Biochemistry 32*, 13019–13025.
101. Hochschwender, S. M., and Laursen, R. A. (1981) The Lysine Binding Sites of Human Plasminogen. Evidence for a Critical Tryptophan in the Binding Site of Kringle 4. *J. Biol. Chem. 256*, 11172–11176.
102. Wu, C. H., Huang, H., Nikolskaya, A., Hu, Z., and Barker, W. C. (2004) The iProClass Integrated Database for Protein Functional Analysis. *Comput. Biol. Chem. 28*, 87–96.
103. Phan, I. Q. H., Pilbout, S. F., Fleischmann, W., and Bairoch, A. (2003) NEWT, a New Taxonomy Portal. *Nucleic Acids Res. 31*, 3822–3823.
104. The UniProt Consortium (2009) The Universal Protein Resource (UniProt) 2009. *Nucleic Acids Res. 37*, D169–D174.
105. Belczewski, A. R., Laplaud, P. M., Chapman, M. J., Marcovina, S. M., and Koschinsky, M. L. (2003) Identification of a Novel Apolipoprotein(a)-Related Protein from the European Hedgehog (*Erinaceus europaeus*). *DNA Sequence 14*, 15–23.
106. Lawn, R. M., Boonmark, N. W., Schwartz, K., Lindahl, G. E., Wade, D. P., Byrne, C. D., Fong, K. J., Meer, K., and Patthy, L. (1995) The Recurring Evolution of Lipoprotein(a). Insights from Cloning of Hedgehog Apolipoprotein(a). *J. Biol. Chem. 270*, 24004–24009.
107. Koschinsky, M. L., Cote, G. P., Gabel, B., and Van der Hoek, Y. Y. (1993) Identification of the Cysteine Residue in Apolipoprotein(a) That Mediates Extracellular Coupling with Apolipoprotein B-100. *J. Biol. Chem. 268*, 19819–19825.
108. Affolter, M., Schaller, J., and Rickli, E. E. (1993) Isolation, Characterization and Partial Amino Acid Sequence of Lamprey Plasminogen. *Protein Sequence Data Anal. 5*, 207–211.
109. Davidson, C. J., Tuddenham, E. G., and McVey, J. H. (2003) 450 Million Years of Hemostasis. *J. Thromb. Haemostasis 1*, 1487–1494.
110. Hughes, A. L. (2000) Modes of Evolution in the Protease and Kringle Domains of the Plasminogen-Prothrombin Family. *Mol. Phylogenet. Evol. 14*, 469–478.
111. Lawn, R. M., Patthy, L., Pesole, G., and Saccone, C. (1997) Apolipoprotein(a): A Puzzling Evolutionary Story. *J. Mol. Evol. 44*, 234–236.

112. Nakamura, Y., Gojobori, T., and Ikemura, T. (2000) Codon Usage Tabulated from International DNA Sequence Databases: Status for the Year 2000. *Nucleic Acids Res. 28*, 292.
113. Ikeo, K., Takahashi, K., and Gojobori, T. (1991) Evolutionary Origin of Numerous Kringles in Human and Simian Apolipoprotein(a). *FEBS Lett. 287*, 146–148.
114. Trexler, M., and Patthy, L. (1983) Folding Autonomy of the Kringle 4 Fragment of Human Plasminogen. *Proc. Natl. Acad. Sci. U.S.A. 80*, 2457–2461.
115. Rejante, M. R., Byeon, I. J., and Llinás, M. (1991) Ligand Specificity of Human Plasminogen Kringle 4. *Biochemistry 30*, 11081–11092.
116. Thomson, A., Cook, A., Meyers, T., and Jenkins, N. (1997) Variation in Glycosylation Heterogeneity of Human Plasminogen Determined by Matrix-Assisted Laser Desorption/Ionization Mass Spectrometry. *Pharm. Sci. 3*, 49–51.
117. Varki, A. (1993) Biological Roles of Oligosaccharides: All of the Theories Are Correct. *Glycobiology 3*, 97–130.
118. Belczewski, A. R., Laplaud, P. M., Chapman, M. J., and Koschinsky, M. L. (1996) The Complete cDNA Sequence Encoding Plasminogen from the European Hedgehog (*Erinaceus europaeus*). *Gene 171*, 271–274.
119. Lawn, R. M., Schwartz, K., and Patthy, L. (1997) Convergent Evolution of Apolipoprotein(a) in Primates and Hedgehog. *Proc. Natl. Acad. Sci. U.S.A. 94*, 11992–11997.
120. Lawn, R. M. (1996) How Often Has Lp(a) Evolved? *Clin. Genet. 49*, 167–174.
121. Rouy, D., Laplaud, P. M., Saboureau, M., and Anglés-Cano, E. (1992) Hedgehog Lipoprotein(a) Is a Modulator of Activation of Plasminogen at the Fibrin Surface. An in Vitro Study. *Arterioscler., Thromb., Vasc. Biol. 12*, 146–154.
122. Battistel, M. D. (2009) *Profiling Plasminogen Kringle Domains According to Ligand Specificity: Structure/Function Correlation*, Ph.D. Dissertation, Carnegie Mellon University, Pittsburgh, PA.
123. Grishaev, A., and Llinás, M. (2005) Protein Structure Elucidation from Minimal NMR Data: The CLOUDS Approach. *Methods Enzymol. 394*, 261–295.
124. Bermejo, G. A., and Llinás, M. (2010) Structure-Oriented Methods for Protein NMR Data Analysis. *Prog. Nucl. Magn. Reson. Spectrosc. 56*, 311–328.