Journal of Advanced Research 24 (2020) 529–543

Contents lists available at ScienceDirect

Journal of Advanced Research

journal homepage: www.elsevier.com/locate/jare

Pharmaceutical modulation of the proteolytic profile of Transforming Growth Factor Beta induced protein (TGFBIp) offers a new avenue for treatment of *TGFBI*-corneal dystrophy

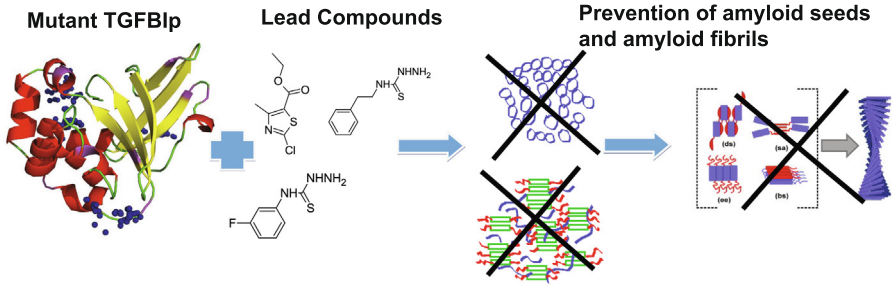
Anandalakshmi Venkatraman [a,1], Minh-Dao Duong-Thi [b,1], Konstantin Pervushin [b,*], Sten Ohlson [b,*], Jodhbir Singh Mehta [c,d,*]

[a] Singapore Eye Research Institute, 11 Third Hospital Avenue, Singapore 168751, Singapore
[b] Nanyang Technological University, School of Biological Sciences, 60 Nanyang Drive, Singapore 637551, Singapore
[c] Singapore National Eye Centre, 11 Third Hospital Avenue, Singapore 168751, Singapore
[d] Ophthalmology and Visual Sciences Academic Clinical Program, Duke-NUS Graduate, Medical School, Singapore 169857, Singapore

HIGHLIGHTS

- Corneal stromal dystrophies are a group of hereditary disorders caused by mutations in the *TGFBI* gene and affect the corneal stroma and epithelium.
- The disease is characterized by the accumulation of insoluble deposits of the mutant TGFBIp leading to poor visual acuity in patients.
- Mutations are hypothesized to disrupt the protein folding and stability, leading oligomerization of the mutant protein.
- Current treatment relies on surgical intervention, either tissue removal or substitution, both of which are associated with disease recurrence.
- The lead compounds reported here prevent/delay the atypical proteolysis of the mutant protein and the generation of amyloidogenic fragments.

GRAPHICAL ABSTRACT



Abbreviations: 1D, 1-Dimensional; 2D, 2-Dimensional; 3D, 3-Dimensional; AA, Amino Acid; BMRB, Biological Magnetic Resonance Data Bank; DMSO, Dimethyl sulfoxide; DSS, 4, 4-dimethyl-4-silapentane-1-sulfonic acid; EIC, Extracted Ion Chromatogram; EMI, Emilin-like domain; FAS1, Fasciclin like Domain; FPLC, Fast Protein Liquid Chromatography; GCD, Granular Corneal Dystrophy; HPLC, High-performance liquid chromatography; HSQC, Heteronuclear Single Quantum Coherence Spectroscopy; ms, Millisecond; MS, Mass spectrometry/spectrometer; IPTG, Isopropyl-beta-D-thiogalactopyranoside; ITC, Isothermal Titration Calorimetry; LB, Luria Bertani; LE, Ligand Efficiency; LCD, Lattice Corneal Dystrophy; MALDI, Matrix-Assisted Laser Desorption/Ionization; PBS, Phosphate Buffered Saline; SD, Standard Deviation; SDS-PAGE, Sodium Dodecyl Sulphate-polyacrylamide gel electrophoresis; SPR, Surface Plasmon Resonance; *TGFBI*, Transforming Growth Factor Beta Induced; TGFBIp, Transforming Growth Factor Beta Induced protein; TOF, Time-of-Flight; TFA, Trifluoroacetic acid; WAC, Weak affinity chromatography; WT, Wild Type.

Peer review under responsibility of Cairo University.

* Corresponding authors at: The Academia, 20 College Road, Discovery Tower Level 6, Singapore 169856, Singapore (J. S. Mehta). School of Biological Sciences, 60 Nanyang Drive, Nanyang Technological University, Singapore 637551, Singapore (S. Ohlson and K. Pervushin).

E-mail addresses: kpervushin@ntu.edu.sg (K. Pervushin), SOhlson@ntu.edu.sg (S. Ohlson), jodmehta@gmail.com (J.S. Mehta).

[1] These authors contributed equally to this work

https://doi.org/10.1016/j.jare.2020.05.012
2090-1232/

ARTICLE INFO

Article history:
Received 11 February 2020
Revised 15 April 2020
Accepted 11 May 2020
Available online 22 May 2020

Keywords:
Corneal dystrophy
Weak affinity chromatography
TGFBIp
Fragment screening
Proteolysis

ABSTRACT

Corneal dystrophies are a group of genetically inherited disorders with mutations in the *TGFBI* gene affecting the Bowman's membrane and the corneal stroma. The mutant TGFBIp is highly aggregation-prone and is deposited in the cornea. Depending on the type of mutation the protein deposits may vary (amyloid, amorphous powdery aggregate or a mixed form of both), making the cornea opaque and thereby decreases visual acuity. The aggregation of the mutant protein is found to be specific with a unique aggregation mechanism distinct to the cornea. The proteolytic processing of the mutant protein is reported to be different compared to the WT protein. The proteolytic processing of mutant protein gives rise to highly amyloidogenic peptide fragments. The current treatment option, available for patients, is tissue replacement surgery that is associated with high recurrence rates. The clinical need for a simple treatment option for corneal dystrophy patients has become highly essential either to prevent the protein aggregation or to dissolve the preformed aggregates. Here, we report the screening of 2500 compounds from the Maybridge RO3 fragment library using weak affinity chromatography (WAC). The primary hits from WAC were validated by ^{15}N-HSQC NMR assays and specific regions of binding were identified. The recombinant mutant proteins (4th FAS-1 domain of R555W and H572R) were subjected to limited proteolysis by trypsin together with the lead compounds identified by NMR assays. The lead compounds (MO07617, RJF00203 and, BTB05094) were effective to delay/prevent the generation of amyloidogenic peptides in the R555W mutant and compounds (RJF00203 and BTB05094) were effective to delay/prevent the generation of amyloidogenic peptides in the H572R mutant. Thus the lead compounds reported here upon further validation and/or modification might be proposed as a potential treatment option to prevent/delay aggregation by inhibiting the formation of amyloidogenic peptides in *TGFBI*-corneal dystrophy.

Introduction:

Protein misfolding and aggregation are hypothesized to play a significant role in many pathological conditions, including Alzheimer's and Parkinson's diseases [1]. Protein aggregation disorders are commonly caused by genetic or epigenetic changes, (physical or chemical) of the primary protein [2]. There are different types of protein aggregation, and the most common type is physical aggregation, where there is a physical association of one protein molecule, with another, without a change in the protein structure. The second process of aggregation is referred to as chemical aggregation, where new chemical bonds are formed. These altered chemical bonds, in the protein, either allow the protein to bind to other interacting partners, or to form chemical bonds with other proteins. This can lead to aggregation or can alter the aggregation properties of the protein [3]. Both aggregation mechanisms may occur simultaneously, depending on the chemical environment, which may produce either a soluble or insoluble protein aggregates [4].

TGFBI-associated corneal dystrophy is a protein aggregation disorder where mutations in the *TGFBI* gene produce a mutant protein (TGFBIp), that is prone to aggregation, and the protein aggregates are deposited in various layers of the cornea [5,6]. The deposition of the protein aggregates leads to corneal clouding and subsequent decrease in visual acuity [7,8]. There are 74 different mutations reported in the *TGFBI* gene [9], and the nature of the amino acid substitution determines the type of protein deposits (e.g. amyloid fibrils, amorphous powdery aggregates or a mixed form); the layer of the cornea in which these proteins are deposited; and the age of onset of the disease [10,11]. TGFBIp is a 68 KDa extracellular protein and is the second most abundant protein in the cornea [12,13]. The majority of the mutations reported in the literature are associated either to the 1st or 4th FAS-1 domains making them mutational hotspots [9,10,14]. Protein deposits in *TGFBI*-associated corneal dystrophies are found only in the cornea, although mutations are observed ubiquitously. Hence there is the possibility of a unique mechanism of aggregation or tissue-specific factors that may play a major role in corneal protein aggregation [15]. It has been hypothesized that there may be a corneal specific factor that may trigger the aggregation cascade and an effective clearance mechanism is absent in the cornea, to process mutant proteins or clear the aggregated proteins.

In a well-folded protein, the hydrophobic residues are not exposed to the external environment (to favor any kind of chemical interaction), thus keeping the protein in a more stable form [16]. In most protein aggregation disorders, it is the partially unfolded, or the misfolded protein (and its intermediates), that serve as precursors for protein aggregation [17]. When a protein becomes partially unfolded, the hydrophobic residues are presented close to the external environment and are the ones responsible for initiating the aggregation process by forming chemical interactions. In most amyloid disorders, it has been postulated that proteins follow a nucleation-dependant aggregation process [18]. The nucleation process is the initial and rate-determining step of the aggregation and amyloid fibril formation. After the nucleation step, there may be the addition of a monomer-cluster aggregation or a cluster–cluster aggregation [19]. The molecules are then built up in an orderly fashion to form insoluble macro-molecules of amyloid fibrils. Hence the amyloid fibrils become complex structures, for any proteolytic processing, and degradation.

There have been several previous studies that have analyzed the differences in the proteolytic processing of the mutant and the WT-TGFBIp [20–24]. Korvatska *et al.* reported the presence of either the full length or shorter truncated TGFBIp, in the amyloid deposits of corneal dystrophy patients with p.Arg124Cys, p.Arg124His and p. Arg124Leu mutations [5]. A study by Stix *et al.* reported the analysis of the composition of amyloid deposits from patients with F540S mutation, the authors observed several small TGFBIp-peptide fragments, which were 6.5 KDa, 6.9 KDa, 14KDa, 17KDa, and 21KDa in addition to the full-length TGFBIp [25]. Recently, proteomic analysis of corneal deposits using Mass Spectrometry (MS) of several amyloidogenic (R124H, V624M, A546D, A546D/P551Q, H626R, and R124C) and non-amyloidogenic mutations (R555W) identified the abundance of 2 peptide sequences TGFBIp 515–533

and TGFBIp 571–588 [20–24] in the corneal aggregates. In our recent MS analysis of corneal deposits from patients with H626R and R124C mutations, we also identified peptides TGFBIp 515–533, TGFBIp 571–588 and TGFBIp 611–642 to be highly enriched in the patient samples compared to the control cornea [11]. The proteomics studies also reported the increased abundance of serine protease HTRA1 (HtrA1) in the amyloid deposits of Lattice Corneal Dystrophy (LCD) patients and proposed the proteolytic processing of mutant protein by HTRA1 [20,26]. All these studies confirm that the proteolytic processing of the mutant protein is different from the WT-TGFBIp. It has also been proposed that the generation of the smaller peptide fragments may act as fibrillation seeds and trigger the aggregation pathway [11]. The mechanism of seeded fibrillation on TGFBIp-A546T mutant protein using an 18 amino acid long amyloidogenic peptide TGFBIp-571–588 has been reported by Andreasen *et al.* [27].

One of the possible strategies to counteract the effect of specific mutations on protein clearance mechanisms might be the development of a low-to-medium-affinity small molecular ligand binding. These ligands bind to mutant protein, modulate the overall thermodynamic stability of the target proteins in biological tissues [28] and prevent the protein from being exposed to uncommon proteolytic processing. The relevance of transient interactions in drug discovery has been a useful guide to identify lead candidates for many diseases that are currently under clinical trial [29,30]. In this context, weak affinity chromatography (WAC) may serve as an important tool to screen for transient interactors to targets that have multiple biological functions. These interactions are effective to create the necessary biological effect and can be used to aim for several targets at the same time, [31–33] for maximum efficiency. WAC has the advantage of high throughput, less utilization of protein targets and fragments, but is a simple yet powerful application using a standard HPLC machine [30,34]. The main challenge for screening weak binders is the difficulty in detecting them and analysing their binding efficiency [30,34,35]. To overcome difficulties in screening for a transient interaction between the compound and TGFBIp, we used WAC combined with MS analysis, adapted to high throughput, to screen [36] for the binding of compounds (2500 fragments from Maybridge library) to WT and mutant TGFBIp.

Here we report three small chemical modulators that can bind to the mutant TGFBIp, and influence its proteolytic processing. The binding of the chemical modulator, to the mutant protein, prevent it from being processed into smaller amyloidogenic peptide fragment seeds, and hence possibly delay the aggregation process. The small molecule compounds were selected from a Maybridge RO3 fragment library of 2500 compounds. We employed WAC technology as the primary transient binding detection tool [35] and verified key compound binding and interaction using ^{15}N-HSQC NMR analysis.

Materials and methods

Chemicals and reagents

PCR reagents, PCR and Gel purification kit, polymerases and restriction enzymes were purchased from Fermentas (Fermentas Inc., Glen Burnie, MD) and Kapa biosystems (Kapabiosystems, Inc., Woburn, MA). pCDF2 vector system with the Ek/LIC cloning kit was purchased from Novagen (Novagen (EMD), Philadelphia, PA) Ni-Sepharose resin was purchased from GE Healthcare (GE Healthcare Life Sciences, Piscataway, NJ). Ampicillin, streptomycin and isopropyl b-D-thiogalactopyranoside (IPTG) were purchased from Sigma- Aldrich (Sigma-Aldrich Inc., MO, USA). LC-MS water and extra pure ammonium acetate were purchased from Merck Millipore (Billerica, MA, USA).

The Maybridge library

Maybridge library RO3 fragment library (RO3-Rule of three) used in the study was purchased from Thermo Fisher Scientific that included 2500 fragments with molecular weights (MW) ranged from 82 to 296 Daltons (Da) and an average mass of 180 Da. ClogP values were from −2.5 to 3.0. Each fragment contained 0–4 rotatable bonds, 0–3 hydrogen bond donors or acceptors. The polar surface area of each molecule was from 0 to 70.04 A^2. Fragments were dissolved in dimethyl sulfoxide (DMSO) as individual compounds at an original concentration of 100 mM. The library was chosen based on the following advantages: The RO3 fragment library offered highly diverse fragments with novel heterocyclic design. The hits generated from the screening are reported to be selective, compact and ligand-efficient leads that may increase success rates with strong ADME (Absorption, Distribution, Metabolism, and Excretion) profiles. The RO3 diversity fragment library offered improved structural diversity profiles and experimental solubility data for the fragments. The fragments in the library were assured to be soluble in both DMSO and PBS buffer (1 nM). The purity of each fragment was verified by NMR analysis and 1-D NMR spectra were available for the compounds.

Expression and purification of WT and mutant TGFBIp

Expression and purification of the 4th FAS-1 domain of WT-TGFBIp (amino acids G^{506} to Q^{633}), mutants R555W and H572R were followed as reported by our group [37]. We used WT-TGFBIp, non-amyloidogenic mutant R555W and amyloidogenic mutant H572R.

Expression and purification of 4th FAS1 WT and mutant TGFBIp proteins

The clones of the WT and mutant proteins were grown in 50 mL LB media at 37° C till an OD of 0.6–0.8 was reached and were induced with 0.5 mM IPTG at 16° C. Following induction, the cultures were grown for 16 h. The cell cultures were examined and verified for protein expression before and after induction before purification of large cultures. Large scale cultures were done in 2 or 4 L of LB broth containing streptomycin till an OD of 0.6–0.8 was reached and induced at 16° C with 0.5 mM IPTG. The cultures were centrifuged at 8000 rpm and the cell pellets were frozen overnight at −80° C. The pellets were then suspended in a cold buffer containing 50 mM Tris, 400 mM NaCl and 5% glycerol at pH 7.4. The suspended pellets were sonicated or lysed using a French press at 1000 psi or a combination of both. The lysed cells were then centrifuged at 18,000 rpm for 45 min to obtain soluble protein fractions. The soluble proteins were mixed with the Ni-Sepharose (GE healthcare) resin and incubated for an hour at 4° C. The supernatant-resin mixtures were loaded on to 15 mL purification columns and washed with a cold buffer containing 50 mM Tris, 400 mM NaCl, 5% glycerol and 20 mM Imidazole at pH 7.4. Further washing and elution were done with increasing imidazole gradients up to 200 mM imidazole. The eluted fractions were concentrated by centrifuging at 3000 rpm for 1–2 h in 20 min intervals using Amicon ultrafilters (Merck Millipore, Billerica, MA, USA). The samples were validated using SDS PAGE. The purified fractions were loaded on to a Fast Protein Liquid Chromatography (FPLC) system, gel filtration columns and eluted against a linear gradient of PBS. The purified fractions were stored at −80° C until further use.

The ^{15}N labeled proteins for NMR experiments were prepared and processed in the same way as the un-labeled samples but grown in M9 minimal media with ^{15}N labeled ammonium chloride. For ^{13}C labeled proteins both ammonium chloride and glucose were replaced with isotope-labeled salts.

F-Pocket analysis on WT-TGFBIp and R555W mutant

F-pocket is an open-source pocket detection package based on Voronoi tessellation and alpha spheres built on top of the publicly available package Qhull. The modular source code is organized around a central library of functions, a basis for three main programs: (i) F-pocket, to perform pocket identification, (ii) T-pocket, to organize pocket detection benchmarking on a set of known protein–ligand complexes, and (iii) D-pocket, to collect pocket descriptor values on a set of proteins [38]. Inputs for the F-pocket analysis were the protein structures of the 4th FAS-1 domain of TGFBIp WT (2LTB) and mutant R555W (2LTC) available from the Protein Data Bank (PDB) [39]. These were the published solution structures of protein determined by NMR analysis, and the PyMOL analysis of overlaying the WT and R555W did not show a major change in the overall secondary structure of the protein except local changes around the mutations [39].

Screening of compounds using weak affinity chromatography (WAC)

The proteins were immobilized on spherical porous diol silica particles (Kromasil, from EKA, Bohus, Sweden, at 5 μm in diameter, 300 Å pore size, ~ 100 m^2 surface area per gram) by reductive amination reaction (Schiff base method). First, the diol silica (~7 mg for each protein) was sonicated in 0.3 mL isopropanol for 2 min and then washed by centrifugation (2000 × g in 2 min) with 4 × 0.3 mL water followed by another 2 min sonication. Periodic acid was added in the slurry to a concentration of 100 mg/mL to oxidize diols on silica into aldehyde groups. The slurry was rotated at room temperature (22° C) for 3 h before being washed with 4 × 0.3 mL water then 0.3 mL of 100 mM sodium phosphate buffer, pH 7.5 (immobilization buffer) for WT and R555W proteins or 100 mM sodium phosphate buffer, pH 7.0 for H572R protein. Resulting aldehyde silica batches were let to react with proteins in PBS containing 10 mg/mL $NaCNBH_3$ (final concentration). Reaction mixtures were rotated at 4° C for 92 h (WT and H572R) or 96 h (R555W) before they were stopped by centrifugation (2000 × g in 2 min) and removal of supernatants, and then washed with 6 × 0.3 mL immobilization buffer. Supernatants and washings were pooled, which were later measured absorbance at 280 nm (NanoDrop 1000, Thermo Fisher Scientific, Wilmington, DE, USA) to calculate the number of immobilized proteins.

Another batch of diol silica was treated the same way except no protein was added into the immobilization reaction mixture. This batch served as blank silica (produced blank column 2015–04-02). Each resulting silica batch was packed into a stainless steel capillary (50 × 0.5 mm) by an air-driven liquid pump (Haskel MS-110, Burbank, CA, USA) at 340 bar for 2 h. The packing mobile phase was 100 mM sodium phosphate buffer pH 7.5. Obtained columns were WT, R555W, H572R and blank columns, packed with corresponding silica batches. Immobilization and packing were carried out shortly before the primary screening to keep the proteins fresh. If not in use, the affinity columns were filled with PBS pH 7.4 and stored at 4°C.

For primary screening, the entire Maybridge fragment library was distributed into 50 mixtures, each comprised of 50 fragments with non-overlapping molecular weight. The mixtures were diluted 100 times with water to give a final concentration of 20 μM for each fragment in 1% DMSO. In the confirmation single solution study, the samples were prepared by diluting original fragments 1000 times with water to a final concentration of 0.1 mM in 0.1% DMSO.

Primary screening of mixtures using WAC

Screening by WAC was performed on a 1200 Agilent High-performance liquid chromatography (HPLC) system that coupled to a 6530 Accurate-Mass quadrupole time-of-flight mass spectrometer, which employed an Agilent Jet Stream Thermal Focusing electrospray ionization source (Agilent Technologies, Santa Clara, CA, USA) (LC-Q-tof). The mobile phase flow rate was 20 μL/min and the injection volume was 0.4 μL. The mobile phase was 10 mM ammonium acetate, pH 6.6. Room temperature was controlled at 18–19 °C during the analyses. MS experiments were performed in single MS, positive or negative ionization mode. Fragmentor voltage was set at 100 V; capillary voltage (Vcap) at 1800 V, drying nitrogen gas was at 12 L/min at 250 °C and nebulizer pressure at 40 psig. Sheath gas flow was at 12 L/min; 300 °C. Skimmer voltage was at 65 V. A reference mixture solution containing ions *m/z* 121.050873 and 922.00979 in positive mode was used to correct the masses in real-time. Chromatograms were acquired in the mass range *m/z* 70 – 930.

The LC-Q-tof was used for the screening experiment and MassHunter B02.01 was used to acquire data; data analyses were performed using MassHunter B06.00 (both versions were from Agilent Technologies). Fragment chromatograms were retrieved by extracted ion chromatograms (EIC) function with their accurate masses, using symmetric expansion at ± 0.5 *m/z*. The extracted chromatogram was smoothed and integrated to obtain time at peak apex, which was considered as the retention time. All injections were duplicated and the analysis time was 45 min for each injection. First, the full Maybridge library was screened by injecting the above 50- fragment mixtures onto the target columns. Primary active retention times (t_{active}) of fragments from this experiment were used to calculate primary dissociation constant (K_d), and subsequently, primary ligand efficiency (LE) based on equation (1) and (2), respectively.

$$K_d = B_{tot}/(t_{active} \times F) \quad (1)$$

$$LE = -(RTlnK_d)/NHA \quad (2)$$

In which B_{tot} is the total number of active sites in the R555W column, assuming to be 50% of immobilized protein amount; t_{active} is the net retention time (t'_R) of a fragment on active column minus net retention time of that fragment on the blank column. Net retention time is retention time minus the void time for DMSO (ion *m/z* 79.0212 in positive mode). Negative t_{active} from the subtraction was normalized to zero. F is the mobile phase flow rate. LE is ligand efficiency in $kcalmol^{-1}$ per atom, R is the universal gas constant ($R = 1.99 \times 10^{-3}$ $kcalK^{-1}mol^{-1}$), T is the absolute temperature in Kelvin (K), NHA is the number of heavy atoms in the fragment. In the primary hit selection, only fragments with highest LE for each target (55 fragments with LE ≥ 0.38 $kcalmol^{-1}$ per heavy atom for WT, 53 fragments with LE ≥ 0.43 kcal mol^{-1} per heavy (non-hydrogen) atom for H572R, and 50 fragments with LE ≥ 0.46 $kcalmol^{-1}$ per heavy atom for R555W) were chosen to enter single solution confirmation step in consideration of our WAC and NMR hit confirmation capacity. In this report, a fragment is considered as a binder when its t_{active} is equal to or higher than ten times of average SD of retention times of fragments on the column.

Primary hit confirmation as single fragments using WAC

The primary hits were then run as singles to confirm detection. The LC-Q-tof conditions were kept the same as above, except the new room temperature was 25–28 °C. Acquisition time and polarity mode to analyse each fragment were set according to information from the mixture screening. The standard deviation (SD) values of retention times of all detected fragments including DMSO on duplicate injections in mixtures were calculated; the average SD on each column was used to set a limit of affinity detection on that column. Fragments were confirmed as hits when their LE $\geq$ 0.30 kcalmol^{-1} per heavy atom and t_{active} were higher than ten times the average SD.

Sample preparation for NMR experimentation

The top 10 compounds that gave the highest LE values for the R555W mutant protein in WAC screening were used for secondary validation by NMR assays. Uniformly ^{15}N-labeled proteins (R555W) were used at concentrations of 0.1 to 0.2 mM proteins in PBS buffer. The NMR samples consisted of 90% H_2O, 10% D_2O, and 5% DSS as an internal standard and transferred to 5 mm NMR tubes (Norell). All NMR spectra were acquired using 600 or 800 MHz Bruker Avance II spectrometers. Chemical shifts were referenced to DSS directly for 1H and indirectly for ^{13}C and ^{15}N spins. The NMR data were processed using TopSpin 2.0 (www.bruker-biospin.com) and analyzed using CARA (cara.nmr.ch).

^{15}N-HSQC experiments giving a correlation between amide nitrogen and proton pairs were acquired as fingerprint spectra for the protein. The mutant R555W protein was then titrated with various concentrations of DMSO). ^{15}N-HSQC experiments were then recorded again at different fragment concentrations (1:1, 1:5 and 1:10) for the top 10 fragments identified by WAC experiments. The interactions of the proteins to the ligand were measured by peak shifts or changes in peak intensity when the ^{15}N-HSQC spectrum of the free protein was overlaid with spectra of DMSO and various fragment concentrations. The NMR assignments for the 4th-FAS1 domain of R555W mutant protein (2LTC) spectra were obtained from the BMRB website and the HSQC spectrum was calibrated accordingly. The peak shifts were analyzed and the amino acid residues that caused the shifts were mapped on to the structure of the protein.

The ^{15}N-HSQC spectrum of 1:10 protein to drug concentration was used for calculating peak shifts. Chemical shift perturbations along the proton and nitrogen dimensions for each amino acid of mutant R555W protein alone were subtracted from the chemical shift perturbations of protein with the compound. The chemical shift values obtained in terms of ppm values were converted to Hertz by multiplying the perturbations with the magnetic strength of the NMR machine used. Root Mean Square Deviation (RMSD) values were also calculated based on the below-mentioned formula and used for heat map generation of chemical shift changes noted in each amino acid of the TGFBIp-R555W mutant protein.

RMSD = Square root [(difference in chemical shifts along proton dimension in Hertz) ^{2} + (difference in chemical shifts along nitrogen dimension in Hertz) ^{2}]

Limited proteolysis by trypsin

The purified 4th FAS-1 domain proteins (50 μg) were subjected to proteolysis by trypsin for 1 h. The proteins were pre-treated with the 3 lead compounds from NMR analysis (MO07617, RJF00203, BTB05094) for about 30 min at room temperature. Samples of 250 μg of each of the proteins were subjected to trypsin digest (1:500) protein-trypsin ratio at 37C. A 50 μg of protein sample was withdrawn after 15 min, 30 min, 45 min and 1 h and the proteolytic reaction was stopped by the addition of formic acid. The samples were air-dried and stored at −20C before MALDI analysis. The samples were suspended in 0.1% TFA /50% acetonitrile solution and spotted on to MALDI matrices and peptide fragments generated by trypsin digest were further analyzed by MALDI-TOF mass spectrometry.

Results

Expression and purification of 4th FAS-1_WT and mutant TGFBIp

The proteins required for all the assays were purified and the correct protein sequence and mass of the proteins were verified by MS analysis. MALDI MS was used to confirm the identity of the purified mutants. The mass of the WT protein showed a single intact species at *m*/*z* ratio of 16,666 ± 10 Da. The mass of the R555W protein showed a single intact species at *m*/*z* r ratio of 16,696 ± 10 Da, and this was in agreement with the predicted molecular mass of the mutant R555W, and the mass of the H572R protein showed a single intact species at *m*/*z* ratio of 16,686 ± 10 Da, and this was in agreement with the predicted molecular mass of the mutant H572R (Supplementary Figure 1A). The pure protein was analyzed using SDS-PAGE that resulted in pure protein as indicated in Supplementary Figure 1B.

F-Pocket analysis of 4th FAS-1 domain of TGFBIp-WT and mutant R555W

The predictions from the F-pocket analysis identified 3 potential binding pockets for WT protein and 4 potential binding pockets for R555W protein (Supplementary Figure 2). Pocket 1 of the WT protein had a druggability score of 0.934 (between scales of 0 to 1), indicating that it is very likely for a compound/fragment to bind to the pocket. The other 2 pockets predicted for the WT did not have a favourable druggability score. For the 4th FAS1 domain of the R555W mutant, the F-pocket analysis identified 4 potential binding pockets with a druggability score of 0.907 for pocket 1. Both the WT and the R555W mutant shared a similar pocket that gave the highest druggability score (Table 1).

Screening of compounds using WAC

Among 2500 fragments of the Maybridge library that were screened on the 4th FAS-1 domain of TGFBIp-WT, R555W, H572R, and blank columns, >88% were detected in each column (Table 2). The remaining fragments were not detectable, either due to their poor ionization in the MS, excessive forming of adduct ions that resulted in new *m*/*z* values and therefore escaped detection, or because they were bound so tightly to the column that their retention time exceeded the monitoring window of 90 min (duplicate injections of 45 min each). Some fragments were detected only once in the duplicate due to a weak signal, which caused missing SD values. The interference of contaminating ions, especially from DMSO, was not considered in the counting. Note that contaminating ions from DMSO only interfere with the detection of non-retained fragments that are co-eluted with DMSO. Results (Table 2) showed that the obtained affinity columns could reliably detect binding in the range of μM to mM. The detection capacity of the columns covered more than two orders of magnitude. A combination of taking into account both the LE and binding affinities, a list of about 134 fragments were selected as initial hits. Since the initial screening was done in the form of 50 compounds in one mixture, the compound that gave the highest affinity score

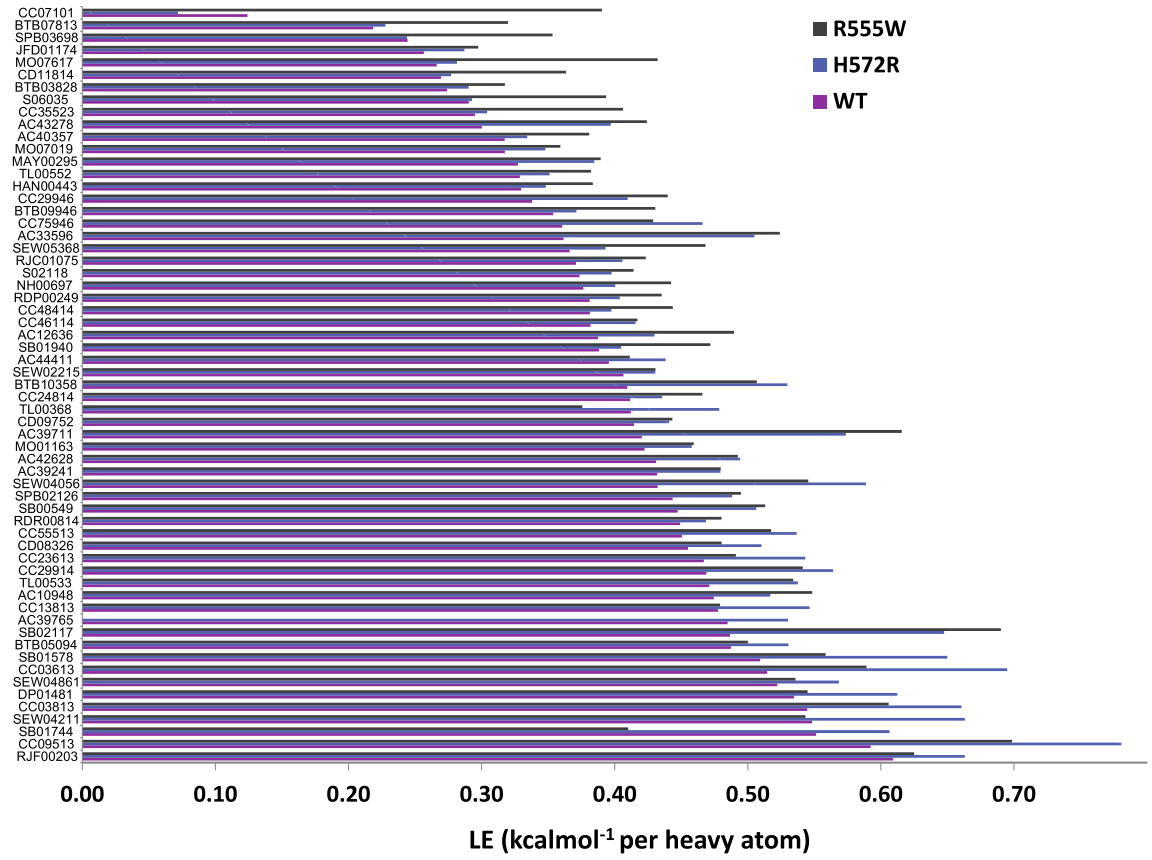


Fig. 1. Ligand Efficiency (LE) values of interactions between the final hits and targets R555W, H572R and WT.

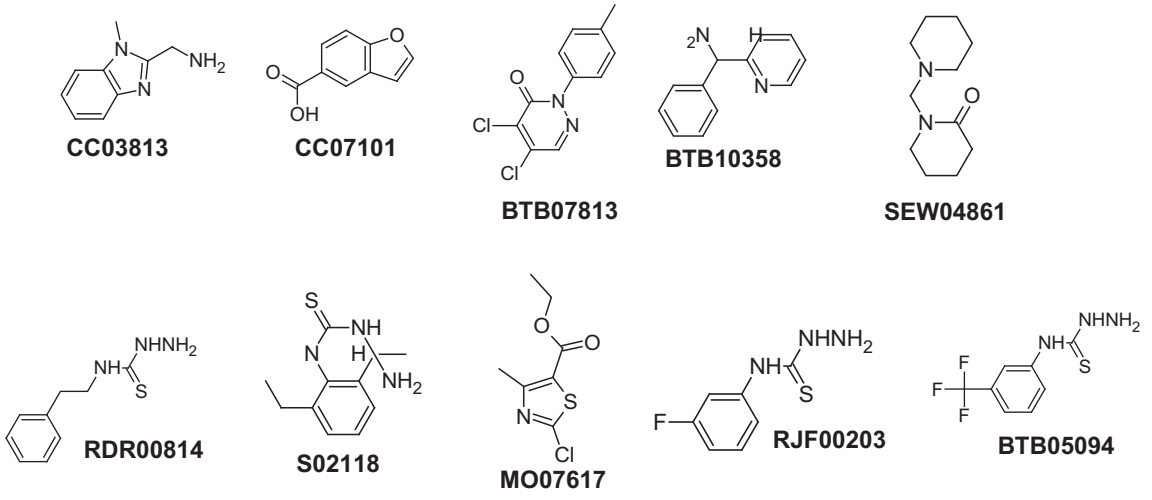


Fig. 2. The list of top 10 compounds that showed higher affinities towards both the 4th FAS-1 domain of TGFBIp-R555W and H572R mutant protein.

was also tested in the form of single compounds to confirm the binding of the compounds to the proteins.

Among 134 fragments that were analyzed as a single solution, 110 fragments were detected on a blank column, 105 fragments on R555W and H572R, and 104 fragments on WT. The discrepancy indicates some fragments bound strongly to the target protein so that they were not eluted, or the signals were weak and unstable in the mass spectrometer. A few fragments, which had been suspected as tight binders, showed up when their analysis times were extended. However, there were still some fragments that did not appear after 300 min on protein columns even though they showed strong signals with the blank column. These fragments were positive hits and, therefore, included in the final list for further validations. It has to be noted that only ~ 50 fragments that had the highest LE in the primary mixture screening were selected for the confirmation (Fig. 1). The constraint in number was due to the confirmation capacity of WAC and NMR in our labs. The retention times of profiles of compounds and DMSO with the column immobilized with proteins of the 4th FAS-1 domain of TGFBIp H572R, WT, and R555W determined by WAC are shown in Fig. 3.

Table 1
Output from the F-Pocket analysis for 4th FAS-1 domain of TGFBIp-WT and mutant R555W.

4th FAS1 domain WT		4th FAS1 domain R555W	
Pocket 1	Pocket score = 30.751 Druggability score = 0.934	Pocket 1	Pocket score = 33.981 Druggability score = 0.907
Pocket 2	Pocket score = 24.9220 Druggability score = 0.0782	Pocket 2	Pocket score = 24.6410 Druggability score = 0.0495
Pocket 3	Pocket score = 18.0877 Druggability score = 0.0505	Pocket 3	Pocket score = 23.9584 Druggability score = 0.0124
		Pocket 4	Pocket score = 23.2564 Druggability score = 0.6165

The initial screening by WAC showed 10 compounds that had a high-affinity binding towards both the mutants and the structure of the compounds are given in Fig. 2 and the binding efficiency values and LE are given in table 3.

Selection of compounds for secondary screening and ^{15}N-HSQC-based NMR validation

Due to the limited capacity to handle a large number of compounds screening by ^{15}N-HSQC-based NMR analysis, the top 10 compounds that showed favourable binding to R555W mutant protein and compounds that had high ligand efficiency towards R555W mutant were chosen for secondary screening by NMR. There were no well-resolved ^{15}N-HSQC spectra available for the WT-TGFBIp and H572R mutant due to the poor stability of proteins and the greater aggregation propensity of the proteins. The proteins aggregated so quickly that by the time a ^{15}N-HSQC spectrum could be recorded there was not enough soluble protein for the spectra to be resolved well without peak broadening (Supplementary Figure 3). Due to the non-availability of ^{15}N-HSQC spectra for the WT and H572R protein to observe chemical shift perturbations and to study the possible drug binding mechanism, the H572R mutant and WT-TGFBIp were not available for the secondary NMR analysis. For ^{15}N-HSQC-based secondary validation, R555W mutant protein was used. Although the 3-D structure of the 4th FAS-1 domain of WT-TGFBIp is available from PDB (protein code 2LTB) determined by solution NMR, the published constructs (aminoacid 502–634) and our designed constructs (aminoacid 506–633) were different [39] and hence could contribute to the differences in the NMR spectra.

Out of the 10 compounds analyzed by NMR, compounds MO07617, RJF00203, and BTB05094 showed greater chemical shift perturbations around the binding pockets as predicted by F-pocket assays. Fig. 4A-C shows the chemical shifts for each of the amino acids along the proton dimension and also along the nitrogen dimension. The amino acid changes are also marked in the 3D structure of the mutant R555W protein as in Fig. 4A. For compound MO07617, the 1:10 protein to compound ^{15}N-HSQC chemical shift perturbations of the individual amino acids shows strong binding of the compound/ligand to the protein as shown in Fig. 4A. Any change in the chemical shifts above + 10 Hz or −10 Hz was identified as positive binding, and the amino acids may be involved in an interaction of the compound/ligand with the protein, as the local interaction of the compound with the amino acids causes huge changes in the chemical shifts. The chemical shift changes along the proton dimension may be due to the interaction through a hydrogen bond. The changes may also be caused by the ring-current effect of the aromatic side chain of compounds. The chemical shifts also coincide with the predicted binding pocket of the protein. Amino acids 545–550 of the R555W mutant protein are near the binding pocket, and the chemical shift changes may be due to the compound/ligand binding to the pocket. For compound RJF00203 and BTB05094 (Fig. 4B and C), there are strong chemical shift perturbations observed around the binding pocket near amino acids 545 to 550. The magnitude of the chemical shifts for the top three compounds near the binding pocket was quite comparable. Since a similar effect was found in all three compounds, it is possible that these compounds interact strongly with the mutant protein. The other seven compounds (CC03813, CC07101, BTB07813, BTB10358, SEW04861, RDR00814, and S02118) showed some chemical shift perturbations that were not as intense as the first 3 compounds and did not show preferential binding near the binding pockets (Supplementary Figure 4).

The Root Mean Square Deviation (RMSD) values for each of the amino acids on how the chemical shift perturbations were extended along the proton and nitrogen dimension were calculated for all the 10 compounds as shown in Fig. 5. Since the protein had 132 amino acids, the RMSD graphs were split into 2 parts with the first 60 amino acids in the first and the rest of the amino acids in the second part. The white color index shows the maximum chemical shift perturbations of individual amino acids. A color index with black color shows no changes or very little changes in that particular amino acid when the compound/ligand is added. The grey to white color index shows greater chemical shift perturbations for that particular amino acid residue. The RMSD values also conclude that the compounds MO07617, RJF00203, and BTB05094 showed strong binding near the predicted binding pocket regions. The concentration of protein titrated to the lead fragments reported in the manuscript is 1:10 for the NMR titration experiment. A concentration of either 0.1 mM or 0.2 mM of the protein was titrated to a 1 mM or 2 mM of the fragment. The titration experiments also showed effective binding at 1:5 concentrations for the lead compounds MO07617, RJF00203 and BTB05094. The optimal dose of treatment needs further experimentation and validation of the lead compounds if it has to be used on animal models/humans. The lead compounds (small fragments) could also be chemically modified/extended to improve efficacy in humans.

The 4 compounds that showed a common motif among the top 10 compounds (Fig. 2) that were analyzed using ^{15}N-HSQC assays, only 3 compounds with that motif showed strong binding to the R555W mutant protein. Out of the 3 compounds, 2 compounds (RJF00203 and BTB05094) showed very strong chemical shift perturbations along the binding pocket region. The other compounds with a similar motif (RDR00814, and S02118) did not have significant chemical shift perturbations along the binding pocket region. The compounds that gave a positive chemical shift response are the ones also having a highly electronegative fluorine molecule in their chemical structure. The fluorine molecules might contribute to the strong binding of the ligand/compound to the mutant R555W protein.

Table 2
SDs of retention times of fragments by duplicate mixture injections and detection of mixture screening on affinity columns.

Column	WT	R555W	H572R	Blank
Average SD (including DMSO) (min)	0.0530 (n = 2299)	0.0584 (n = 2337)	0.0425 (n = 2243)	0.0402 (n = 2393)
Highest value of SD (min)	1.415	1.596	2.351	1.477
Number of detected fragments	2252 (90%)	2298 (92%)	2199 (88%)	2345 (94%)
K_d range detected (μM)	19–2629	13 – 1732	9 – 1101	Not applicable

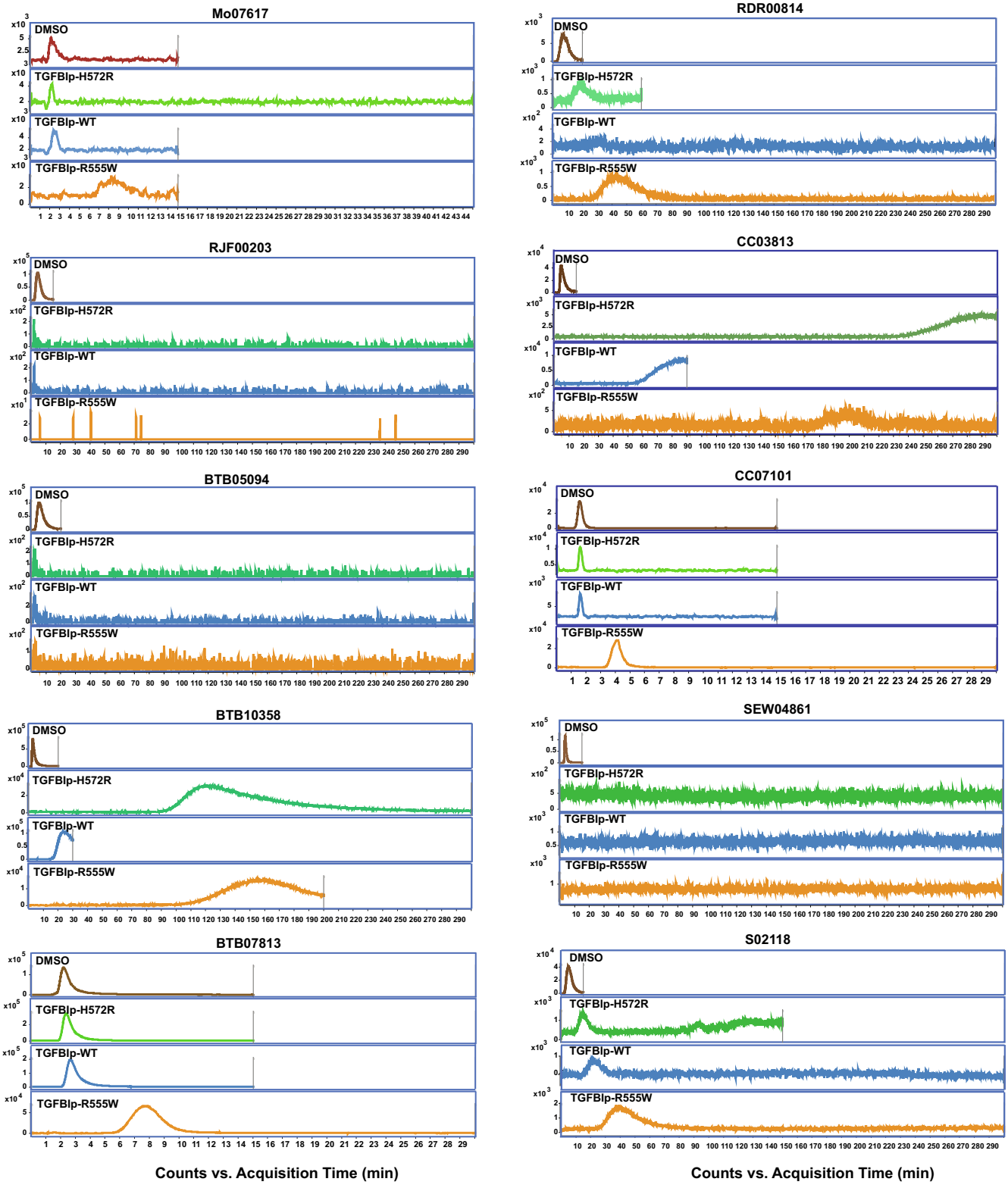


Fig. 3. Retention time profiles of compounds and DMSO with the column immobilised with proteins of the 4th FAS-1 domain of TGFBIp H572R, WT and R555W.

Table 3
List of compounds with their binding efficiency and LE used for secondary NMR screening.

No	CODE	K_d on R555W (μM)	K_d on H572R (μM)	K_d on WT (μM)	LE on R555W (kcalmol^{-1} per heavy atom)	LE on H572R (kcalmol^{-1} per heavy atom)	LE on WT (kcalmol^{-1} per heavy atom)
1	CC03813	5	2	17	0.61	0.66	0.54
2	CC07101	382	No binding	No binding	0.39	No binding	No binding
3	BTB07813	184	No binding	No binding	0.32	No binding	No binding
4	BTB10358	7	4	66	0.51	0.53	0.41
5	SEW04861	< 3	< 2	< 5	> 0.54	> 0.57	> 0.52
6	RDR00814	28	36	55	0.48	0.47	0.45
7	S02118	29	44	82	0.41	0.40	0.37
8	MO07617	164	No binding	No binding	0.43	No binding	No binding
9	RJF00203	< 3	< 2	< 5	> 0.63	> 0.66	> 0.61
10	BTB05094	< 3	< 2	< 5	> 0.50	> 0.53	> 0.49

Limited proteolysis of 4th FAS-1 domain of TGFBIp-WT, R555W and H572R proteins

The 4th FAS-1 domain of TGFBIp-WT and the mutants were subjected to limited proteolysis by using trypsin enzyme with and without the three lead compounds (MO07617, RJF00203, and BTB05094) from drug screening and the peptide patterns generated by trypsin enzyme digest were analyzed at different time points (Table 4 and 5). The R555W mutant protein, without compounds, was subjected to proteolysis by trypsin, 15 min of proteolysis produced 5 peptides that could be detected by MALDI-MS. After 30 min, 45 min and 1 h, while the number of peptides remained the same, there was a production of the highly amyloidogenic peptide sequence (E^{611}PVAEPDIMATNGVVHVITNVL^{632}) after 30 min of the enzyme digest. The observed peptide sequence is known to be highly amyloidogenic as predicted by various algorithms, and our group [11] has studied the aggregation properties of this peptide sequence in much detail. The peptide sequence of 11 reported clinical mutations responsible for the amyloid phenotype is located around this region (amino acid 611 to 632) of the protein. When the R555W mutant protein was treated with the lead compounds (MO07617, RJF00203, and BTB05094) for 30 min at room temperature before the trypsin digest, the highly amyloidogenic peptide sequence was not produced. This phenomenon was observed for all the 3 lead compounds. The compounds might have interacted with the exposed hydrophobic amino acids and form chemical interaction involving the residues and protecting the amino acid from trypsin enzymatic digestion. This process of chemical stabilization of the mutant protein with the lead compounds might have prevented the formation of amyloidogenic peptide seeds.

Similarly, the H572 mutant protein was subjected to limited proteolysis by trypsin, and the peptide fragments were analyzed by MALDI-MS. The mutant H572R protein in the absence of any compounds produced 4 different peptides 15 min after enzyme digestion. The longer amyloidogenic peptide (N^{603}NVVSVNKEPVAEPDIMATNGVVHVITNVL^{632}) was observed. The same peptide fragments were observed after 30 min, 45 min and 1 h, and they included the longer amyloidogenic peptide. This peptide is an extended version of the amyloidogenic peptide observed in the R555W mutant (E^{611}PVAEPDIMATNGVVHVITNVL^{632}). When the H572R mutant protein was treated with compound MO07617, there was still a generation of the amyloidogenic peptide, whereas when the protein was treated with the other 2 compounds RJF00203 and BTB05094, the long amyloidogenic fragment was not generated 45 min after the enzyme digest. The compound MO07617 does not bind to the H572R protein according to the initial WAC screening, whereas the other 2 compounds had a higher affinity to the mutant protein. The proteolysis reported here is only limited proteolysis but should be followed for longer time points to understand the proteolytic process in greater detail.

Discussion:

In this drug discovery approach, we identified a possible drug-binding pocket with the highest druggability score for both the 4th FAS-1 domain WT and R555W mutant protein, as the 3D structure was available only for these two proteins. Based on our laboratory experience, certain mutants of the TGFBIp protein is difficult to be expressed and purified from E.coli cells as the protein has high aggregation potential and is highly unstable in solution and may reduce solubility in solution [40]. Due to the high aggregation potential of the mutant proteins, we were unable to run long 2D and 3D NMR experiments to determine the structure of the protein bound to the lead compounds. We are exploring other options like X-ray crystallography and solid-state NMR to determine the structure of the protein-compound bound complex. The transient interaction of the 2500 compounds (50 mixtures with 50 compounds in each mixture with differences in the masses), was screened using WAC against WT and the 2 mutants. The assays confirmed about 61 lead compounds. The binding of the top 10 compounds with the greatest LE and affinity towards the R555W mutant was also verified by ^{15}N-HSQC NMR assays. A well-resolved HSQC spectrum was not available for ^{15}N-HSQC NMR assays for WT and H572R. The protein was not very stable in PBS solution in which all the drug binding ^{15}N-HSQC NMR assays were tested.

The secondary NMR assays using chemical shift perturbations were analyzed for each of the compounds to R555W mutant titration at different compound concentrations. The secondary NMR assays were useful in eliminating false positives produced by WAC screening and identified 3 lead compounds that bound with a greater affinity near the binding pockets predicted by the f-pocket analysis. The 3 lead compounds were added to the protein, limited proteolysis of the protein using a serine protease (trypsin enzyme), was done for a maximum of one hour, and the peptide fragments were analyzed by MALDI-MS. The addition of lead compounds to both the mutants prevented the formation of amyloidogenic peptide sequences, which could trigger the aggregation cascade. All the 3 reported compounds MO07617, RJF00203, and BTB05094 were effective in preventing the production of the highly amyloidogenic peptide when treated with mutant R555W (GCD phenotype) protein, while compounds RJF00203 and BTB05094 were effective in preventing the production of the highly amyloidogenic peptide when treated with mutant H572R (LCD phenotype) protein.

The limited proteolysis analysis was used as a probe to understand the structure of the protein [41]. According to previous studies, only the highly unstructured or the most flexible region of the protein was subjected to proteolysis [42]. In mutant TGFBIp, the presence of a different amino acid in place of the native amino acids might have caused these regions to be unstructured. The mutations may have also induced a local change in the secondary structure of the protein [39]. In our study, the addition of com-

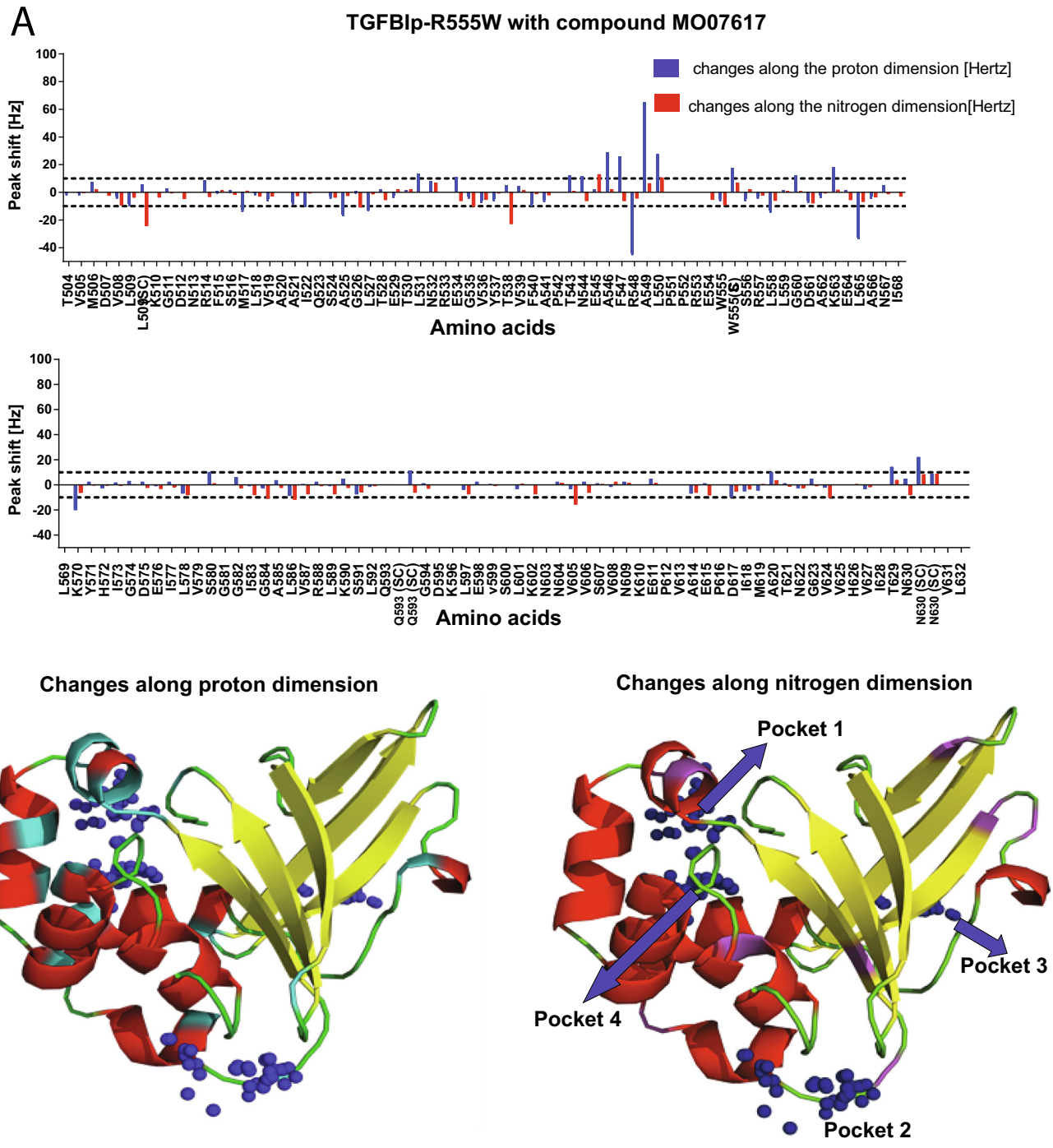


Fig. 4. A to C: Chemical shift perturbations of individual amino acid of 4th FAS-1 domain of TGFBIp with **MO07617** compound (4A) **RJF00203** compound (4B) and **BTB05094** (4C). Chemical shift perturbations that showed a difference of above ± 10 Hertz along the proton dimension marked in cyan on the available 3D structure from PDB (2LTC) of TGFBIp-R555W mutant protein. The blue spheres are the predicted binding pockets by f-pocket analysis. Chemical shift perturbations that showed a difference of above ± 10 Hertz along the nitrogen dimension marked in magenta on the available 3D structure from PDB (2LTC) of TGFBIp-R555W mutant protein. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

pounds to the mutant protein has caused changes in the secondary structure of the protein as verified by the NMR assays. The addition of compounds to the mutant proteins may have stabilized the local secondary structure on the surface of the protein. The compounds may have made new chemical bonds or interactions with amino acids along with the binding pocket, thereby not providing access to the protease to act on the protein. This may have prevented the formation of amyloid peptides with high aggregation propensity that could, in turn, act as amyloid fibril seeds to accelerate amyloid fibril formation or protein aggregation. The mutation-induced changes in TGFBIp may also play a role in the normal turnover rate of the protein [43]. The phenotypic expression of each of the mutations may be directly affected by the site of mutation on TGFBIp which in turn governs the proteolytic processing and the

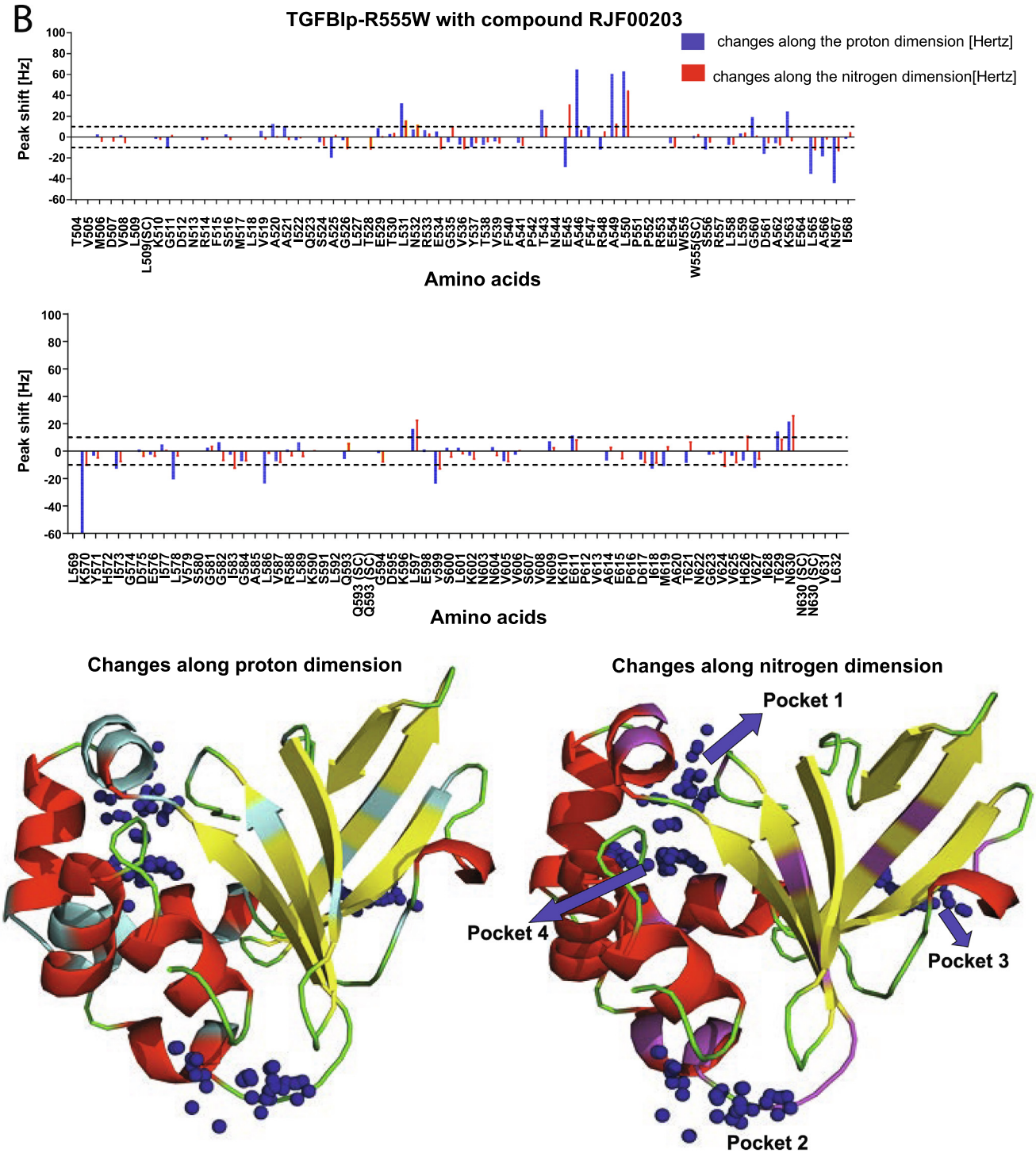


Fig. 4 (*continued*)

turnover rate of the protein [43]. In the mutant protein, there is a higher exposure of these regions to proteolysis activity giving rise to highly amyloidogenic peptides.

There have been previous reports on the abnormal proteolysis processing of the mutant TGFBIp as a pivotal step in protein aggregation and amyloid deposition in corneal dystrophy [5,20,22–24]. The addition of compounds to the mutant proteins has stabilized the local secondary structure on the surface of the protein. These lead compounds reported in this study acted directly on the proteolytic processing of the mutant proteins. The compounds may have made new chemical bonds or interactions with amino acids along with the binding pocket, thereby not providing access to the protease to act on the protein. This prevented the formation of amyloid peptides with high aggregation propensity that could, in turn, act as amyloid fibril seeds to accelerate amyloid fibril formation or protein aggregation. The two lead compounds RJF00203 and

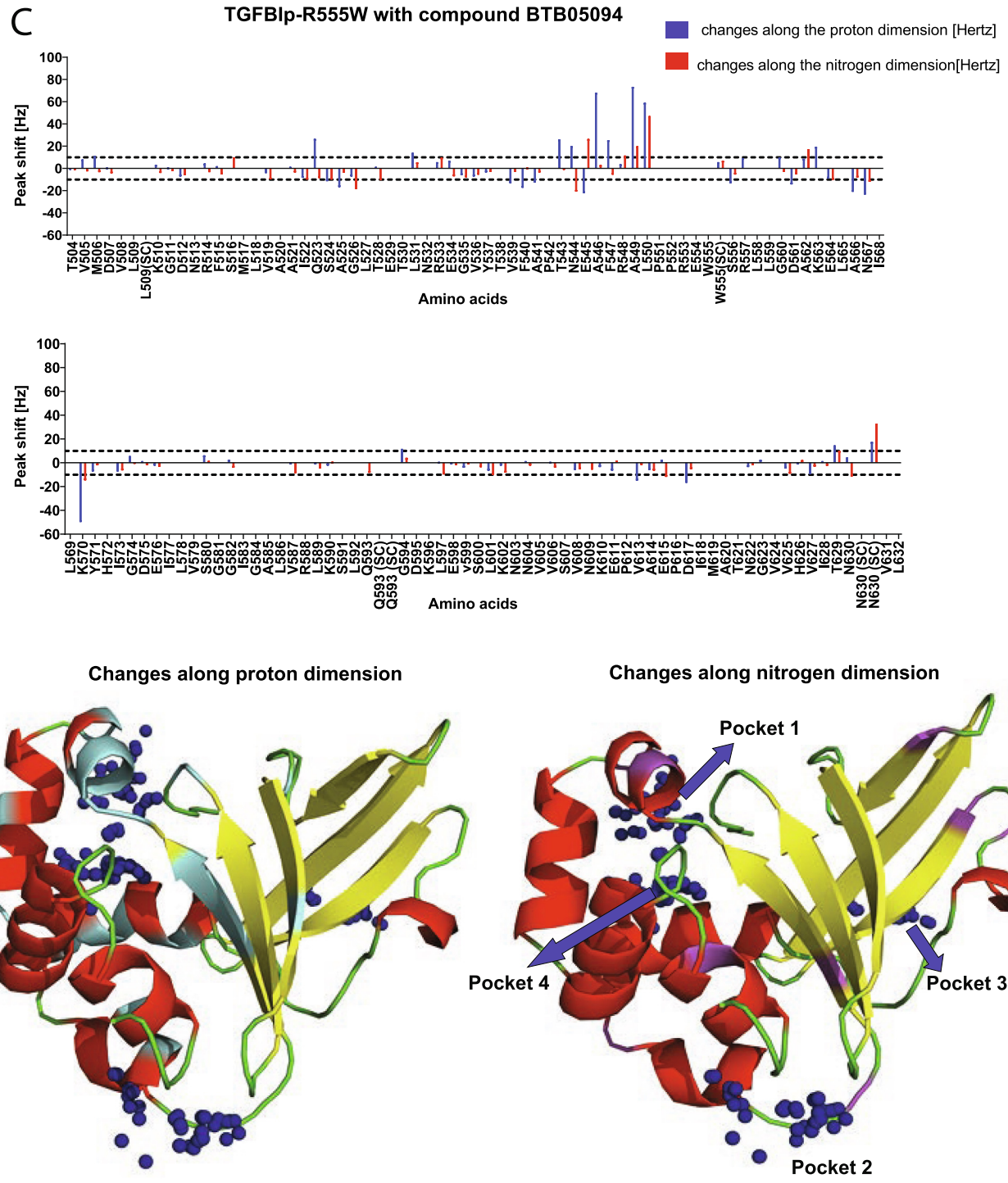


Fig. 4 (*continued*)

BTB05094 reported here may be explored as direct treatment options to delay/prevent protein aggregation in *TGFBI*-associated corneal dystrophies.

In summary, the lead compounds reported here were tested against both LCD mutation (H572R) and GCD mutation (R555W). The compounds were effective against the production of amyloidogenic peptides that may trigger aggregation in corneal dystrophy. These compounds acted directly or with a chemical modification may be used to modulate the affinity of TGFBIp and thus delay protein aggregation in corneal dystrophies.

Conclusion and future Direction:

TGFBI-corneal dystrophy seems to have a complex phenotype and heterogenic expression. There has been no reliable animal model available for these particular mutations to test the efficacy of the compounds. The available animal model to represent the disease is a transgenic mouse for R124H by Yamazoe K *et al* [44]. We have previously examined the corneal deposits in these animals, and we identified that the expression of TGFBIp was very much lower in the animals with R124H mutation compared to

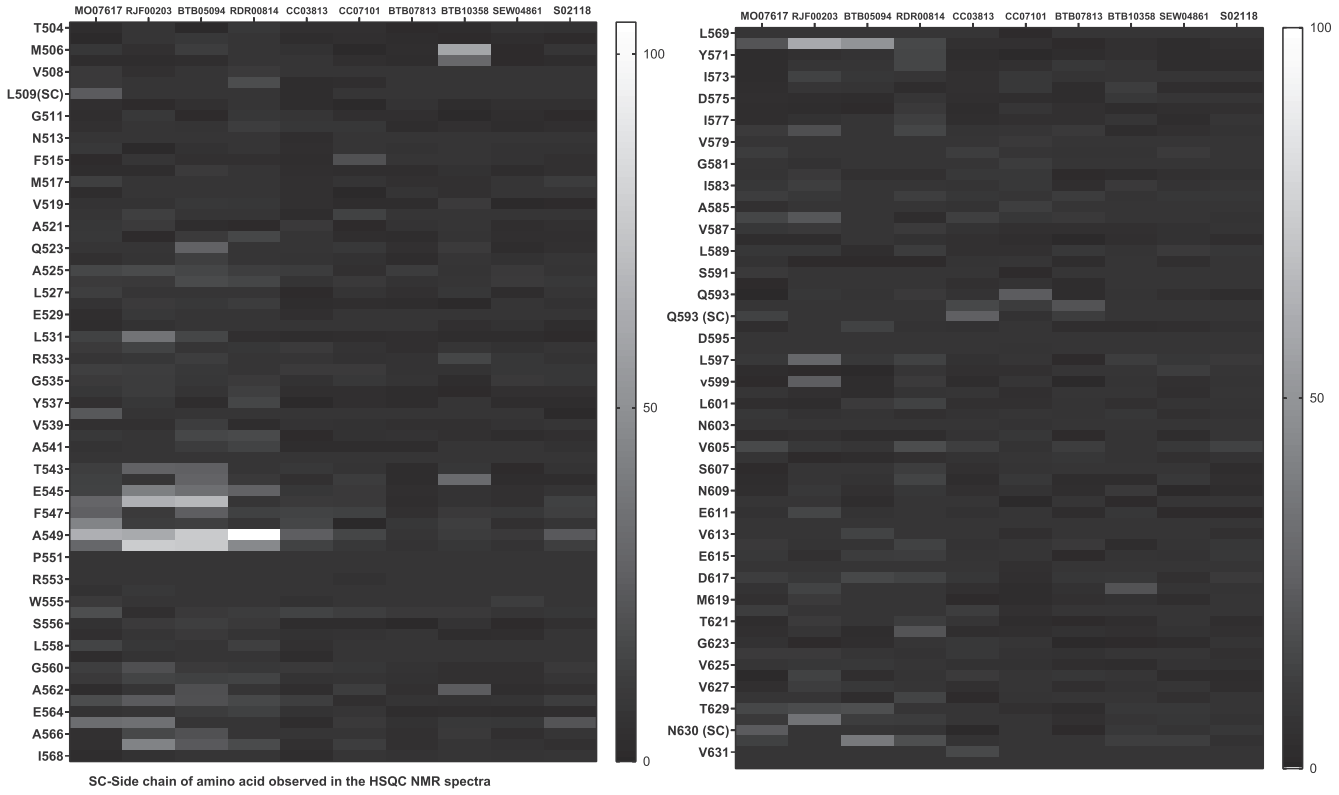


Fig. 5. RMSD values for the 132 amino acid of TGFBIp-R555W mutant protein to top 10 compound binding validated by ^{15}N-HSQC based NMR assays.

the WT mouse (Results not published). There is a very recent publication "Generation of a mouse model of *TGFBI*-R124C corneal dystrophy using CRISPR/Cas9-mediated homology-directed repair" by Kitamoto *et al* [45] and our group will analyse the feasibility of using these animal models in the future to validate the lead compounds.

To the best of our knowledge, this is the first study to report High throughput screening of a large compound library against the mutant TGFBIp. This study may serve as the initial screening effort to find effective inhibitors to delay or prevent mutant TGFBIp aggregation. Even the other reported compounds that had higher binding affinities in the WAC screening may be effective against other mutant forms of TGFBIp. In the future, these compounds may also be tested with other mutations in the 4th FAS-1 domain of TGFBIp. The lead compounds may also be chemically enhanced to increase binding affinity for better performance. Our group also aims to validate lead compound binding specificity by Surface Plasmon Resonance (SPR) and Isothermal Titration Calorimetry (ITC) experiments of the mutant protein to lead compounds. Once validated for toxicity, bioavailability, mode of action and efficacy using an in-vitro/in-vivo model system the compounds can be a simple but efficacious treatment option available to *TGFBI*-corneal dystrophy patients.

Acknowledgments

The authors would like to acknowledge the funding support from SNEC-HREF R1076/91/2013, R1429/12/2017 and SERI Lee –Foundation Pilot Grant R1586/85/2018. The authors would like to thank Dr. Gary Peh Swee Lim, Ms. Khadijah Binte Adnan, Ms. Dawn Neo Jing Hui and Ms. Jagjit Kaur for helping with the experiments.

Competing Interest

All authors declare no competing interests.

Author Contribution

VA and MDT contributed equally to the manuscript. VA and MDT conceived, planned, carried out the experiments, analyzed results and wrote the manuscript in consultation with SO, KP and JSM. VA and KP performed NMR experiments and analysed results. MDT and SO performed WAC experiments. VA and JSM conceived the study and were in charge of overall direction and planning. All authors provided critical feedback, and helped shape the research, analysis and manuscript.

Table 4
Peptide fragments generated by limited proteolysis of R555W with trypsin and analysis by MALDI mass fingerprinting. The highlighted peptide sequence is the highly amyloidogenic sequence appearing in the mutant protein but absent in the mutant protein with the lead compounds.

	15 min	30 min	45 min	1 h
R555W No compound	SLQGDKLEVSLK	EGVYTVFAPTNEAFR	EGVYTVFAPTNEAFR	ELANILK
	EGVYTVFAPTNEAFR	YHIGDEILVSGGIGALVR	YHIGDEILVSGGIGALVR	EGVYTVFAPTNEAFR
	YHIGDEILVSGGIGALVR	FSMLVAAIQSAGLTETLNR	FSMLVAAIQSAGLTETLNR	YHIGDEILVSGGIGALVR
	FSMLVAAIQSAGLTETLNR	**EPVAEPDIMATNGVVHVITNVL**	**EPVAEPDIMATNGVVHVITNVL**	FSMLVAAIQSAGLTETLNR
	NNVVSVNKEPVAEPDIMATNGVVHVITNVL	NNVVSVNKEPVAEPDIMATNGVVHVITNVL	NNVVSVNKEPVAEPDIMATNGVVHVITNVL	**EPVAEPDIMATNGVVHVITNVL**
R555W with MO07617	ALPPREWSR	EGVYTVFAPTNEAFR	EGVYTVFAPTNEAFR	EGVYTVFAPTNEAFR
	EGVYTVFAPTNEAFR	YHIGDEILVSGGIGALVR	YHIGDEILVSGGIGALVR	YHIGDEILVSGGIGALVR
	YHIGDEILVSGGIGALVR	FSMLVAAIQSAGLTETLNR	FSMLVAAIQSAGLTETLNR	FSMLVAAIQSAGLTETLNR
	FSMLVAAIQSAGLTETLNR	NNVVSVNKEPVAEPDIMATNGVVHVITNVL	NNVVSVNKEPVAEPDIMATNGVVHVITNVL	NNVVSVNKEPVAEPDIMATNGVVHVITNVL
	NNVVSVNKEPVAEPDIMATNGVVHVITNVL			
R555W with RJF00203	EGVYTVFAPTNEAFR	EGVYTVFAPTNEAFR	EGVYTVFAPTNEAFR	EGVYTVFAPTNEAFR
	YHIGDEILVSGGIGALVR	YHIGDEILVSGGIGALVR	YHIGDEILVSGGIGALVR	YHIGDEILVSGGIGALVR
	FSMLVAAIQSAGLTETLNR	FSMLVAAIQSAGLTETLNR	FSMLVAAIQSAGLTETLNR	FSMLVAAIQSAGLTETLNR
	NNVVSVNKEPVAEPDIMATNGVVHVITNVL	NNVVSVNKEPVAEPDIMATNGVVHVITNVL	NNVVSVNKEPVAEPDIMATNGVVHVITNVL	NNVVSVNKEPVAEPDIMATNGVVHVITNVL
R555W with BTB05094	EGVYTVFAPTNEAFR	EGVYTVFAPTNEAFR	SLQGDKLEVSLK	ELANILK
	YHIGDEILVSGGIGALVR	YHIGDEILVSGGIGALVR	EGVYTVFAPTNEAFR	SLQGDKLEVSLK
	FSMLVAAIQSAGLTETLNR	FSMLVAAIQSAGLTETLNR	YHIGDEILVSGGIGALVR	EGVYTVFAPTNEAFR
	NNVVSVNKEPVAEPDIMATNGVVHVITNVL	NNVVSVNKEPVAEPDIMATNGVVHVITNVL	FSMLVAAIQSAGLTETLNR	YHIGDEILVSGGIGALVR
			NNVVSVNKEPVAEPDIMATNGVVHVITNVL	FSMLVAAIQSAGLTETLNR

Table 5
Peptide fragments generated by limited proteolysis of H572R with trypsin and analysis by MALDI mass fingerprinting. The highlighted peptide sequence is the highly amyloidogenic sequence appearing in the mutant protein but absent in the mutant protein with the lead compounds.

	15 min	30 min	45 min	1 h
H572R No compound	IGDEILVSGGIGALVR	IGDEILVSGGIGALVR	IGDEILVSGGIGALVR	IGDEILVSGGIGALVR
	EGVYTVFAPTNEAFR	EGVYTVFAPTNEAFR	EGVYTVFAPTNEAFR	EGVYTVFAPTNEAFR
	FSMLVAAIQSAGLTETLNR	FSMLVAAIQSAGLTETLNR	FSMLVAAIQSAGLTETLNR	FSMLVAAIQSAGLTETLNR
	NNVVSVNKEPVAEPDIMATNGVVHVITNVL	NNVVSVNKEPVAEPDIMATNGVVHVITNVL	NNVVSVNKEPVAEPDIMATNGVVHVITNVL	NNVVSVNKEPVAEPDIMATNGVVHVITNVL
H572R with MO07617	IGDEILVSGGIGALVR	IGDEILVSGGIGALVR	IGDEILVSGGIGALVR	IGDEILVSGGIGALVR
	EGVYTVFAPTNEAFR	EGVYTVFAPTNEAFR	EGVYTVFAPTNEAFR	EGVYTVFAPTNEAFR
	FSMLVAAIQSAGLTETLNR	FSMLVAAIQSAGLTETLNR	FSMLVAAIQSAGLTETLNR	FSMLVAAIQSAGLTETLNR
	NNVVSVNKEPVAEPDIMATNGVVHVITNVL	NNVVSVNKEPVAEPDIMATNGVVHVITNVL		NNVVSVNKEPVAEPDIMATNGVVHVITNVL
H572R with RJF00203	IGDEILVSGGIGALVR	IGDEILVSGGIGALVR	IGDEILVSGGIGALVR	IGDEILVSGGIGALVR
	EGVYTVFAPTNEAFR	EGVYTVFAPTNEAFR	EGVYTVFAPTNEAFR	EGVYTVFAPTNEAFR
	FSMLVAAIQSAGLTETLNR	FSMLVAAIQSAGLTETLNR	FSMLVAAIQSAGLTETLNR	FSMLVAAIQSAGLTETLNR
	NNVVSVNKEPVAEPDIMATNGVVHVITNVL	NNVVSVNKEPVAEPDIMATNGVVHVITNVL		
H572R with BTB05094	IGDEILVSGGIGALVR	IGDEILVSGGIGALVR	IGDEILVSGGIGALVR	IGDEILVSGGIGALVR
	EGVYTVFAPTNEAFR	EGVYTVFAPTNEAFR	EGVYTVFAPTNEAFR	EGVYTVFAPTNEAFR
	FSMLVAAIQSAGLTETLNR	FSMLVAAIQSAGLTETLNR	FSMLVAAIQSAGLTETLNR	FSMLVAAIQSAGLTETLNR
	NNVVSVNKEPVAEPDIMATNGVVHVITNVL	NNVVSVNKEPVAEPDIMATNGVVHVITNVL		

PDB codes used in the manuscript.
2LTB and 2LTC are the PDB codes used in the manuscript.
List of Bioinformatics tools used in the manuscript.

F-pocket analysis (http://fpocket.sourceforge.net/run_online.html)
Cara (http://www.cara.nmr.ch/)
PyMol (https://pymol.org/)
BMRB (http://www.bmrb.wisc.edu/)
Protein Data Bank (PDB) https://www.rcsb.org/search

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.jare.2020.05.012.

References

[1] 1. Stefani M. Protein misfolding and aggregation: new examples in medicine and biology of the dark side of the protein world. Biochimica et Biophysica Acta (BBA) - Molecular Basis of Disease. 2004;1739(1):5-25.
[2] Koo EH, Lansbury Jr PT, Kelly JW. Amyloid diseases: abnormal protein aggregation in neurodegeneration. Proc Natl Acad Sci U S A. 1999;96(18):9989–90.
[3] Finke JM, Roy M, Zimm BH, Jennings PA. Aggregation events occur prior to stable intermediate formation during refolding of interleukin 1beta. Biochemistry 2000;39(3):575–83.
[4] Wang W. Protein aggregation and its inhibition in biopharmaceutics. Int J Pharm 2005;289(1–2):1–30.
[5] Korvatska E, Henry H, Mashima Y, Yamada M, Bachmann C, Munier FL, et al. Amyloid and Non-amyloid Forms of 5q31-linked Corneal Dystrophy Resulting from Kerato-epithelin Mutations at Arg-124 Are Associated with Abnormal Turnover of the Protein. J Biol Chem 2000;275(15):11465–9.
[6] Karring H, Runager K, Valnickova Z, Thøgersen IB, Møller-Pedersen T, Klintworth GK, et al. Differential expression and processing of transforming growth factor beta induced protein (TGFBIp) in the normal human cornea during postnatal development and aging. Exp Eye Res 2010;90(1):57–62.
[7] Weiss JS. Corneal Dystrophies: Molecular Genetics to Therapeutic Intervention—Fifth ARVO/Pfizer Ophthalmics Research Institute Conference. Investigative Opthalmology & Visual Science. 2010;51(11):5391.
[8] Surguchev A, Surguchov A. Conformational diseases: Looking into the eyes. Brain Res Bull 2010;81(1):12–24.
[9] Kheir V, Cortés-González V, Zenteno JC, Schorderet DF. Mutation update: TGFBI pathogenic and likely pathogenic variants in corneal dystrophies. Hum Mutat 2019;40(6):675–93.
[10] Lakshminarayanan R, Chaurasia SS, Anandalakshmi V, Chai SM, Murugan E, Vithana EN, et al. Clinical and genetic aspects of the TGFBI-associated corneal dystrophies. Ocul Surf. 2014;12(4):234–51.
[11] Anandalakshmi V, Murugan E, Leng EGT, Ting LW, Chaurasia SS, Yamazaki T, et al. Effect of position-specific single-point mutations and biophysical characterization of amyloidogenic peptide fragments identified from lattice corneal dystrophy patients. Biochem J 2017;474(10):1705–25.
[12] Klintworth GK, Bao W, Afshari NA. Two mutations in the TGFBI (BIGH3) gene associated with lattice corneal dystrophy in an extensively studied family. Invest Ophthalmol Vis Sci 2004;45(5):1382–8.
[13] Dyrlund TF, Poulsen ET, Scavenius C, Nikolajsen CL, Thogersen IB, Vorum H, et al. Human cornea proteome: identification and quantitation of the proteins of the three main layers including epithelium, stroma, and endothelium. J Proteome Res. 2012;11(8):4231–9.
[14] Kannabiran C, Klintworth GK. TGFBI gene mutations in corneal dystrophies. Hum Mutat. 2006;27(7):615–25.
[15] Thapa N, Lee BH, Kim IS. TGFBIp/betaig-h3 protein: a versatile matrix molecule induced by TGF-beta. The international journal of biochemistry & cell biology. 2007;39(12):2183–94.
[16] Uversky VN, Fink AL. Conformational constraints for amyloid fibrillation: the importance of being unfolded. BBA 2004;1698(2):131–53.
[17] Fink AL. Protein aggregation: folding aggregates, inclusion bodies and amyloid. Fold Des 1998;3(1):R9–R23.
[18] Jarrett JT, Lansbury Jr PT. Seeding "one-dimensional crystallization" of amyloid: a pathogenic mechanism in Alzheimer's disease and scrapie?. Cell 1993;73(6):1055–8.
[19] Speed MA, King J, Wang DI. Polymerization mechanism of polypeptide chain aggregation. Biotechnol Bioeng 1997;54(4):333–43.
[20] Karring H, Poulsen ET, Runager K, Thøgersen IB, Klintworth GK, Højrup P, et al. Serine protease HtrA1 accumulates in corneal transforming growth factor beta induced protein (TGFBIp) amyloid deposits. Mol Vis. 2013;19:861–76.
[21] Karring H, Runager K, Thogersen IB, Klintworth GK, Hojrup P, Enghild JJ. Composition and proteolytic processing of corneal deposits associated with mutations in the TGFBI gene. Exp Eye Res. 2012;96(1):163–70.
[22] Courtney DG, Toftgaard Poulsen E, Kennedy S, Moore JE, Atkinson SD, Maurizi E, et al. Protein Composition of TGFBI-R124C- and TGFBI-R555W- Associated Aggregates Suggests Multiple Mechanisms Leading to Lattice and Granular Corneal Dystrophy. Investigative Opthalmology & Visual Science. 2015;56(8):4653.
[23] Poulsen ET, Runager K, Risor MW, Dyrlund TF, Scavenius C, Karring H, et al. Comparison of two phenotypically distinct lattice corneal dystrophies caused by mutations in the transforming growth factor beta induced (TGFBI) gene. Proteomics Clin Appl. 2014;8(3–4):168–77.
[24] Poulsen ET, Nielsen NS, Jensen MM, Nielsen E, Hjortdal J, Kim EK, et al. LASIK surgery of granular corneal dystrophy type 2 patients leads to accumulation and differential proteolytic processing of transforming growth factor beta-induced protein (TGFBIp). Proteomics 2016;16(3):539–43.
[25] Stix B, Leber M, Bingemer P, Gross C, Ruschoff J, Fandrich M, et al. Hereditary Lattice Corneal Dystrophy Is Associated with Corneal Amyloid Deposits Enclosing C-Terminal Fragments of Keratoepithelin. Investigative Opthalmology & Visual Science. 2005;46(4):1133.
[26] Venkatraman A, Dutta B, Murugan E, Piliang H, Lakshminaryanan R, Sook Yee AC, et al. Proteomic Analysis of Amyloid Corneal Aggregates from TGFBI-H626R Lattice Corneal Dystrophy Patient Implicates Serine-Protease HTRA1 in Mutation-Specific Pathogenesis of TGFBIp. J Proteome Res. 2017;16(8):2899–913.
[27] Andreasen M, Nielsen SB, Runager K, Christiansen G, Nielsen NC, Enghild JJ, et al. Polymorphic fibrillation of the destabilized fourth fasciclin-1 domain mutant A546T of the Transforming growth factor-beta-induced protein (TGFBIp) occurs through multiple pathways with different oligomeric intermediates. The Journal of biological chemistry. 2012;287(41):34730–42.
[28] Martinez Molina D, Jafari R, Ignatushchenko M, Seki T, Larsson EA, Dan C, et al. Monitoring drug target engagement in cells and tissues using the cellular thermal shift assay. Science 2013;341(6141):84–7.
[29] Ohlson S. Designing transient binding drugs: a new concept for drug discovery. Drug Discovery Today 2008;13(9–10):433–9.
[30] Duong-Thi MD, Bergstrom M, Fex T, Svensson S, Ohlson S, Isaksson R. Weak affinity chromatography for evaluation of stereoisomers in early drug discovery. J Biomol Screen 2013;18(6):748–55.
[31] Morphy R, Rankovic Z. Fragments, network biology and designing multiple ligands. Drug Discovery Today 2007;12(3–4):156–60.
[32] Hopkins A, Mason J, Overington J. Can we rationally design promiscuous drugs?. Curr Opin Struct Biol 2006;16(1):127–36.
[33] Frantz S. Drug discovery: Playing dirty. Nature 2005;437(7061):942–3.
[34] Duong-Thi M-D, Bergström M, Fex T, Isaksson R, Ohlson S. High-Throughput Fragment Screening by Affinity LC-MS. J Biomol Screen 2013;18(2):160–71.
[35] Duong-Thi MD, Meiby E, Bergstrom M, Fex T, Isaksson R, Ohlson S. Weak affinity chromatography as a new approach for fragment screening in drug discovery. Anal Biochem 2011;414(1):138–46.
[36] Ohlson S, Shoravi S, Fex T, Isaksson R. Screening for transient biological interactions as applied to albumin ligands: a new concept for drug discovery. Anal Biochem 2006;359(1):120–3.
[37] Elavazhagan M, Lakshminarayanan R, Zhou L, Ting LW, Tong L, Beuerman RW, et al. Expression, purification and characterization of fourth FAS1 domain of TGFβIp-associated corneal dystrophic mutants. Protein Expr Purif 2012;84(1):108–15.
[38] Le Guilloux V, Schmidtke P, Tuffery P. Fpocket: an open source platform for ligand pocket detection. BMC Bioinf 2009;10:168.
[39] Underhaug J, Koldso H, Runager K, Nielsen JT, Sorensen CS, Kristensen T, et al. Mutation in transforming growth factor beta induced protein associated with granular corneal dystrophy type 1 reduces the proteolytic susceptibility through local structural stabilization. BBA 2013;1834(12):2812–22.
[40] Stenvang M, Schafer NP, Malmos KG, Pérez A-MW, Niembro O, Sormanni P, et al. Corneal Dystrophy Mutations Drive Pathogenesis by Targeting TGFBIp Stability and Solubility in a Latent Amyloid-forming Domain. J Mol Biol 2018;430(8):1116–40.
[41] Fontana A, de Laureto PP, Spolaore B, Frare E, Picotti P, Zambonin M. Probing protein structure by limited proteolysis. Acta Biochim Pol 2004;51(2):299–321.
[42] Fontana A, Fassina G, Vita C, Dalzoppo D, Zamai M, Zambonin M. Correlation between sites of limited proteolysis and segmental mobility in thermolysin. Biochemistry 1986;25(8):1847–51.
[43] Poulsen ET, Nielsen NS, Scavenius C, Mogensen EH, Risor MW, Runager K, et al. The serine protease HtrA1 cleaves misfolded transforming growth factor beta-induced protein (TGFBIp) and induces amyloid formation. The Journal of biological chemistry 2019.
[44] Yamazoe K, Yoshida S, Yasuda M, Hatou S, Inagaki E, Ogawa Y, et al. Development of a Transgenic Mouse with R124H Human TGFBI Mutation Associated with Granular Corneal Dystrophy Type 2. PLoS ONE 2015;10(7):e0133397.
[45] Kitamoto K, Taketani Y, Fujii W, Inamochi A, Toyono T, Miyai T, et al. Generation of mouse model of TGFBI-R124C corneal dystrophy using CRISPR/Cas9-mediated homology-directed repair. Sci Rep 2020;10(1):2000.