ChemComm

RSC Publishing

FEATURE ARTICLE

View Article Online
View Journal | View Issue

Cite this: *Chem. Commun.*, 2013, **49**, 4591

Blocking formation of large protein aggregates by small peptides

Yi-Pin Chang[a] and Yen-Ho Chu*[b]

Abnormal protein aggregation is responsible for a variety of human disorders, including Alzheimer's disease, Creutzfeldt–Jakob disease, systemic amyloidosis, and α_1-antitrypsin deficiency (AATD). These diseases are collectively termed conformational diseases and many of them are lethal and have no cure to date. The pathogenesis of these clinical conditions shares a common mechanism of disease, *i.e.*, formation of large protein tangles through intermolecular linkages, cross-β-sheets in particular. These tenacious aggregates are difficult to eradicate and accumulate in cells or tissues over affected individual's lifetime, which eventually leads to devastating consequences. The chronic process is commonly underdiagnosed at an early stage and the prevalence of conformational disease is higher than generally realized. AATD is a typical conformational disease causing both lung and liver disorders, and the World Health Organization (WHO) has highlighted the healthcare problem in 1996. The mechanism of AATD has been unraveled and it serves as an excellent model for the study of conformational disease. Point mutations render α_1-antitrypsin susceptible to self-aggregation and form stacked β-sheets, which is the hallmark of AATD and other conformational diseases. The key to attenuating these diseases is searching for small molecules capable of preventing the formation of large aggregates and dissociating the pre-existing oligomers. To this end, a chemical approach has been developed to tackle the biological problem.

Received 15th October 2012,
Accepted 22nd March 2013

DOI: 10.1039/c3cc37518h

www.rsc.org/chemcomm

[a] *Torrey Pines Institute for Molecular Studies, 11350 SW Village Parkway, Port St. Lucie, FL 34987-2352, USA. E-mail: ychang@tpims.org; Fax: +1 772 345 3649; Tel: +1 772 345 4800*

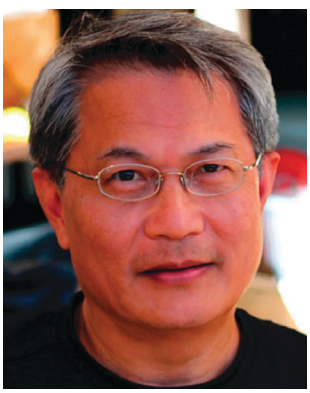
[b] *Department of Chemistry and Biochemistry, National Chung Cheng University, 168 University Road, Minhsiung, Chiayi 62102, Taiwan, ROC. E-mail: cheyhc@ccu.edu.tw; Fax: +886 5 272 1040; Tel: +886 5 242 8148*

Introduction

Proteins self-assemble into ordered structures and play functional roles in living organisms, ranging from spider webs for prey catching to microtubules for cell division.

Yi-Pin Chang

Yi-Pin Chang was born in Taiwan. He received BEng in Chemical Engineering at Feng Chia University, MS in Applied Chemistry at Chinese Culture University and PhD in Chemistry and Biochemistry at National Chung Cheng University in 2007. In collaboration with Dr Ravi Mahadeva at Cambridge University, he commenced multi-disciplinary research in Prof. Yen-Ho Chu's group. He moved to Prof. Benjamin G. Davis' group at Oxford University in 2008, and worked with Prof. Terry D. Butters at Oxford Glycobiology Institute. He joined Torrey Pines Institute for Molecular Studies in 2012 and is doing research on the interface between chemistry and biology.



Yen-Ho Chu

Yen-Ho Chu is a professor of chemistry and biochemistry at National Chung Cheng University in Taiwan. He received his BS in Chemistry at National Tsing Hua University, MS in Biochemical Sciences at National Taiwan University with Professor Kung-Tsung Wang, and PhD in 1993 at Harvard University with Professor George M. Whitesides. He was a postdoctoral fellow with Professor Christopher T. Walsh at Harvard Medical School, a staff scientist with Professor Barry L. Karger at Barnett Institute, and an assistant professor at Ohio State University from 1995 to 1999. His research interests are biomolecular recognition, combinatorial organic synthesis and ionic liquids.

Conversely, protein misfolding and abnormal aggregation can also cause nuisances in molecular biology labs, economic issues in biotechnology industries, or life-threatening diseases in humans. These protein-derived extracellular fibrils or intracellular inclusions accumulate gradually and result in a variety of neurodegenerative and nonneuropathic disorders such as Huntington's disease, Parkinson's disease, thrombosis and type 2 diabetes.[1] Recently, aggregation of the mutant tumor suppressor p53 has also been shown to cause cancers.[2] Apart from these clinical conditions, there is a unique category of disorders derived from the serine protease inhibitor (serpin) superfamily and named serpinopathies.[3] α_1-Antitrypsin deficiency (AATD) is the best-characterized serpinopathies and the only known genetic disorder that leads to chronic obstructive pulmonary disease (COPD). According to the Alpha-1 Foundation, AATD is a hereditary condition which may result in serious lung disease in adults and/or liver disease in infants, children and adults. It has been estimated that approximately 200 000 patients are suffering from the severe type of AATD in the United States and Europe.[4] However, the clinical manifestations of AATD are commonly underdiagnosed or misdiagnosed, causing much morbidity and mortality in those affected.[5] An awareness, screening and detection program sponsored by the Florida Department of Health and Human Services, the Alpha-1 Foundation and the University of Florida has taken place since 1999.

AATD was first observed in young patients with lung emphysema in Sweden,[6,7] and is suspected to be responsible for the premature death of the Polish composer Frédéric François Chopin.[8] The World Health Organization (WHO) suggested the incidence of AATD in Europe and North America is comparable to that of cystic fibrosis, at 1 in 7000 to 1 in 20 000.[9] For the past few decades, the crystal structure, major substrate, inhibitory function, pathogenic mechanism and clinical manifestation of α_1-antitrypsin (AT) have been revealed. However, there is no cure or amiable treatment for AATD patients currently, only supplemental treatments temporarily alleviate the symptoms and organ transplantation is the last resort. Like many other conformational diseases, pathogenic polymerization of the protein not only induces cellular damage by large aggregates, but also results in loss of protein function itself. In a nutshell, blocking abnormal protein aggregation is a crucial strategy for AATD and other conformational diseases. The challenge lies in several aspects of chemistry and biology and must be taken into account for the design and screening of anti-protein aggregation ligands. Firstly, the conserved tertiary structure of the protein superfamily (serpin in the case of AT) and the nearly identical protein sequences between native and pathogenic proteins (point mutation) are the two main obstacles. Secondly, the extraordinarily slow binding kinetics of the ligand possibly due to the active form of the protein is in a thermodynamically metastable state. Thirdly, a conformation-sensitive screening platform allows high-throughput screening of a mixture-based library. Finally, determination of whether the identified binding ligand could retain protein function, facilitate degradation or simply block aggregation.

Small peptides represent a good starting point for the search of anti-protein aggregation ligands when looking into the mechanism of disease and the potential binding sites. Considering the Anfinsen's dogma and the Levinthal's paradox, small peptides with preferential residues may act similar to molecular chaperons to stabilize the native-like conformation or disassemble misfolded aggregates by competing with the normal millisecond folding. In addition, structural fluctuation of metastable proteins and misfolded conformers may expose hydrophobic residues accidentally that are generally buried in the native structure, and thus short flexible peptides are good candidates to maneuver to the binding site. The interactions between residues of protein and small peptide may undergo a process of trial and error to find its lowest-energy structure. Binding kinetics is likely to be one important factor that determines the fate of the binding complex. Prolonged association of small peptides is in principle advantageous to block pathogenic aggregation and may trigger the endoplasmic-reticulum-associated protein degradation and thus facilitate the clearance of misfolded protein.[10] On the other hand, fast dissociation might be beneficial for gain-of-function of secretory proteins and thus restore protein homeostasis. The key questions are how to identify the correct combination of the residues on the small peptide and the design of a peptide library based on the mechanism of disease.

The advances in combinatorial chemistry have made preparation of a large library feasible and thus the main consideration is how to find the needle in the haystack. Therefore, a conformation-sensitive assay based on conventional gel electrophoresis was developed and it allows high-throughput screening of a mixture-based library for the search of anti-protein aggregation ligands. This assay does not require an expensive instrumentation setup and no labeling or chromophoric reagents are needed. Through exquisite design and systematical screening of libraries from libraries, small peptides were identified to selectively block the formation of large aggregates of the pathogenic protein. The most potent peptide has also been used as a chemical tool to validate the classic loop–sheet polymerization model of serpin. In addition, the ability to intervene and reverse polymerize as well as to promote secretion of AT *in vivo* has been demonstrated most recently. Based on the shared mechanism of disease, the developed chemical approach is promising and can be applied to other conformational diseases. This feature article presents how chemistry faces a biological problem underlying a large number and diverse human diseases and focuses on: (1) how biochemical assays are developed based on the mechanism of conformational disease; (2) how small focused libraries facilitate the design, screening and validation of the mixture-based peptide library; (3) characterization of the combinatorially selected ligands; and will begin with (4) the molecular basis of the conformational disease.

Mechanism of α_1-antitrypsin deficiency and conformational disease

AATD is an excellent model for the interrogation of conformational disease, as the β-strand promiscuity of mutant AT is the

shared root of pathogenesis.[11] The normal M-AT is the most abundant plasma protein synthesized in the liver, but function as a major protease inhibitor within the lungs. The disease mechanism has a close connection to the inhibitory function of AT. This 52 kDa glycoprotein serves as a suicide substrate to inhibit its cognate enzyme neutrophil elastase (NE) and thus protects the connective tissues from destruction.[12] A single-nucleotide mutation in the gene (14q32.1) such as the resulting S (Gly264Val) and Z (Glu342Lys) alleles can cause liver deposition and hence plasma deficiency, and manifest cirrhosis and emphysema respectively.[13] It has been estimated that at least 116 million carriers are with the mild combination alleles of MS and MZ, while 3.4 million individuals are with the SS, SZ and ZZ severe alleles worldwide.[14] Despite a relatively low similarity in the primary structure, the tertiary structure of serpins exhibits great homology and share a conserved core domain, which includes the three dominant β-sheets (A, B and C) and the nine α-helices (A through I).[15] In the native state of inhibitory serpins such as AT (Fig. 1), a unique reactive center loop (RCL) is exposed to entrap its targeted protease NE. Following the cleavage of the scissile bond P1–P1′ (Met358-Ser359), the mobile RCL inserts between strands 3 and 5 of the β-sheet A as an additional strand 4 (s4A). This event concomitantly translocates NE more than 70 Å to the opposite end of the molecule with the formation of a 1 : 1 enzyme : inhibitor complex to inactivate the protease, and renders the β-sheet A fully antiparallel.[16] The most striking features of the serpin inhibitory mechanism are the dramatic conformational change and suicidal and irreversible nature of inhibition, in which the RCL plays a crucial role but also comes at price.

The pathogenic aggregates of AT are derived from loop–sheet polymerization.[17] In spite of the pivotal role of the RCL in inhibitory function, point mutations may perturb the protein structure and render the sheet vulnerable to loop insertion by the RCL of adjacent AT at the s4A position. Moreover, the exposed RCL and the unoccupied sheet on the resulting dimer are available for further cross-linkages by other AT molecules to form oligomers and large aggregates eventually. For example, the disruption of salt bridge Glu342-Lys290 by the Z mutation not only extends the RCL as a donor but also expands the β-sheet as a receptor for polymerization.[18] This spontaneous and uncontrolled loop–sheet polymerization is the molecular basis of AATD-associated liver and lung diseases. The same mechanism of diseases is not limited to AT, other pathogenic polymerization of serpins has also been identified including α_1-antichymotrypsin (ACT; Leu55Pro and Pro228Ala), antithrombin (ATIII; Pro54Thr and Asn158Asp), C1 inhibitor (Phe52Ser, Pro54Leu, Ala349Thr, Val366Met, Phe370Ser and Pro391Ser) and neuroserpin (Ser49Pro) in association with emphysema, thrombosis, angio-oedema and dementia (familial encephalopathy with neuroserpin inclusion bodies; FENIB), respectively.[19–22]

RCL peptides bind to serpins

The RCL of serpins has distinct conformational characteristics and accounts for serpin function. The structure of the serpin RCL varies, such as the helical loop in ovalbumin,[23] the β-strand in the latent form of plasminogen activator inhibitor-1 (PAI-1) and antithrombin,[24,25] and the distorted helix in α_1-antichymotrypsin (ACT).[26] Apart from the stable and canonical structure, the RCL of AT can adopt a β-strand conformation and is the basis of pathogenic polymerization.[27] In retrospect, quite a few RCL-derived peptides (AT, ATIII and PAI-1) were synthesized to study the biochemical properties of serpins (Table 1; some peptides from ref. 30, 32 and 38 were exogenous or combinatorially selected). Regardless of the purposes of these studies, RCL peptides generally have poor affinity and selectivity to serpins. Nonetheless, these RCL peptides inspired the peptide annealing method as shown in Fig. 1.[28] Given that the inhibitory mechanism of serpin is conducted by the intramolecular incorporation of the RCL, it appears that RCL segments are able to bind to the s4A site and thus intervene with the propagation of aggregation.

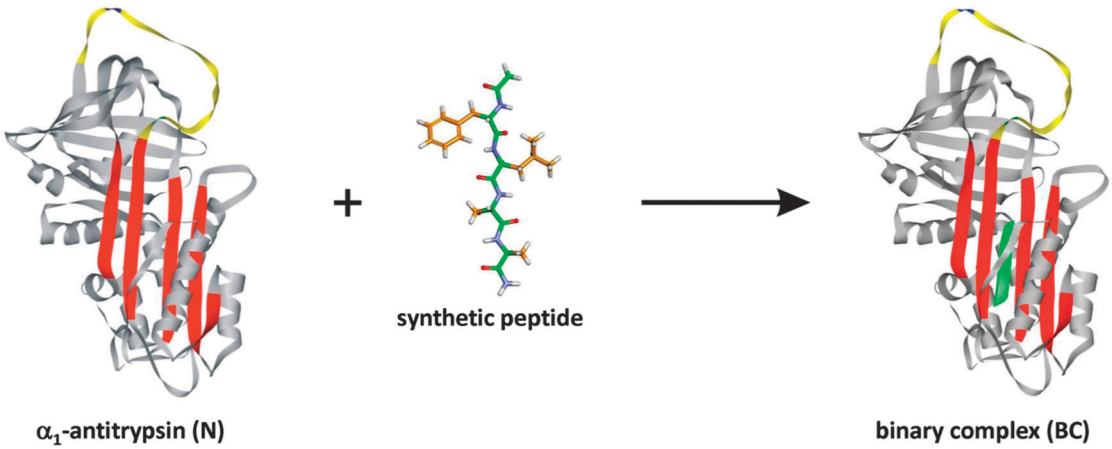


Fig. 1 Targeting the β-sheet A of α_1-antitrypsin to block the pathogenic polymerization. The native form (N) of α_1-antitrypsin (PDB code: 1QLP) contains a reactive center loop (yellow) and a β-sheet A (red) which are responsible for both the function and pathogenesis of α_1-antitrypsin. The five strands in the β-sheet are named as s1A, s2A, s3A, s5A and s6A (from left to right). The synthetic peptide (light green) inserts into the s4A position to form the binary complex (BC) and thus blocks the initiation of loop–sheet polymerization.

Table 1 Serpin-binding peptides derived from RCL, exogenous and combinatorially selected sequences

Peptide sequence[a] Length/origin[b]	Binding serpin[b]	Molar ratio of serpin : peptide[c]	Incubation time and temperature[c]	Ref.
Ac-TEAAGAMFLEAIVM-OH 14-mer/AT RCL homologue	M-AT	1 : 100; 1 : 200	Up to 48 h	40–42
Ac-SEAAASTAVVIAGR-OH 14-mer/ATIII RCL	ATIII	1 : 93	Up to 30 h	43
Ac-SEAAASTAVVIAG-OH 13-mer/ATIII RCL	ATIII M and Z-AT	1 : 100	48 h Up to 144 h (41 °C)	18, 44 and 45
Ac-SEAAAS-OH[d] 6-mer/ATIII RCL	ATIII	1 : 100	Up to 48 h	44 and 46
GTEAAGAMFLEAIPMY 16-mer/AT RCL homologue	M-AT	1 : 200	15 h (48 °C)	47
MFLEAIPMC 9-mer/AT RCL homologue	M-AT	1 : 100	0.5 h (23 °C)	47
Ac-TVASSSTAVIVSAR 14-mer/PAI-1 RCL	PAI-1	1 : 79	Up to 2 h	48
Ac-TVASSSTA 8-mer/PAI-1 RCL	PAI-1	1 : 50	22 h	49
Ac-TEASSSTA 8-mer/PAI-1 RCL homologue	PAI-1	1 : 250	22 h	49
AC-GTEAAGAMFLEAI-OH 13-mer/AT RCL	M-AT	1 : 100	Up to 24 h	50
Ac-TVASS-NH_2[e] 5-mer/PAI-1 RCL	PAI-1 (Ala335Glu)	1 : 4	Until completion	51
Ac-SEAAASTAVVIA-OH 12-mer/ATIII RCL	ATIII M and Z-AT ACT	1 : 75; 1 : 100 1 : 100 1 : 50	Up to 48 h Up to 120 h Up to 16 h	29, 30, 50 and 52–54
Ac-AAGAMFLEAIVM-OH 12-mer/AT RCL	M-AT	1 : 200	Until completion	42
Formyl-AAGAMFLEAIVM-OH 12-mer/AT RCL	M-AT	1 : 200	Until completion	42
Ac-AGAMFLEAIVM-OH 11-mer/AT RCL	M-AT	1 : 200	Until completion	42
Ac-GAMFLEAIVM-OH 10-mer/AT RCL	M-AT	1 : 200	Until completion	42
Ac-AMFLEAIVM-OH 9-mer/AT RCL	M-AT	1 : 200	Until completion	42
Ac-MFLEAIVM-OH 8-mer/AT RCL	M-AT	1 : 200	Until completion	42
Ac-TEAAGAMFLEA-OH 11-mer/AT RCL	M-AT	1 : 200	Until completion	42
Ac-TEAAGAM-OH 7-mer/AT RCL	M-AT	1 : 200	Until completion	42
Ac-FLEAIG-OH 6-mer/AT RCL homologue	Z-AT AT (His334Ala) AT (His334Ser)	1 : 100	Up to 72 h	29, 30, 38 and 54
Ac-SEAAAST-OH[d] 7-mer/ATIII RCL	ATIII	1 : 100	Up to 24 h	30, 46 and 54
Ac-TAVVIA-OH 6-mer/ATIII RCL	M-AT	1 : 100	Up to 48 h	30
Formyl-NGTFVG-OH 6-mer/ACT delta form	M-AT	1 : 200	48 h	30
Ac-AVVIA-OH 5-mer/ATIII RCL	ATIII	1 : 100	24 h	30
Ac-TFLEA-OH 5-mer/AT RCL homologue	M-AT	1 : 200	48 h	30
Formyl-TFVVI-OH 5-mer/ATIII RCL	M and Z-AT	1 : 200	48 h	30
Formyl-FVVII-OH 5-mer/ATIII RCL	M and Z-AT	1 : 200	48 h	30
Formyl-TFVV-OH 4-mer/ATIII RCL	M-AT > Z-AT	1 : 200	48 h	30
Formyl-FVVI-OH 4-mer/ATIII RCL	M and Z-AT	1 : 200	48 h	30
Formyl-VVII-OH 4-mer/ATIII RCL	M and Z-AT	1 : 200	48 h	30
WMDF-NH_2[d] 4-mer/cholecystokinin	ATIII, M-AT	1 : 100	Up to 24 h	30
FMRF-NH_2 4-mer/cholecystokinin	ATIII	Data not shown	Data not shown	30
FLRF-NH_2 4-mer/cholecystokinin	ATIII	Data not shown	Data not shown	30

Table 1 (*continued*)

Peptide sequence[a] Length/origin[b]	Binding serpin[b]	Molar ratio of serpin : peptide[c]	Incubation time and temperature[c]	Ref.
Formyl-MLF-OH[d] 3-mer/commercial source	ATIII	1 : 100	48 h	30 and 54
Formyl-Nle-LF-OH[d] 3-mer/commercial source	ATIII	1 : 100	48 h	54
Ac-FAEAIG-OH 6-mer/alanine-scanning library	M-AT	1 : 100	72 h	32
Ac-FLAAIG-OH 6-mer/alanine-scanning library	M and Z-AT	1 : 100	72 h	32
Ac-FLEAAG-OH 6-mer/alanine-scanning library	M and Z-AT	1 : 100	72 h	32
Ac-FLEAIA-OH 6-mer/alanine-scanning library	Z-AT	1 : 100	72 h	32
Ac-FLEAA-NH_2 5-mer/truncation library	Z-AT	1 : 25	Up to 72 h	32
Ac-LEAAG-NH_2 5-mer/truncation library	M and Z-AT	1 : 100	72 h	32
Ac-LAAIG-NH_2 5-mer/truncation library	M-AT	1 : 100	72 h	32
Ac-FLAA-NH_2 4-mer/truncation library	M and Z-AT	1 : 25	Up to 72 h	32
Ac-AAIG-NH_2 4-mer/truncation library	M-AT	1 : 100	72 h	32
Ac-FLADA-NH_2 4-mer/D-amino acid scanning library	M-AT	1 : 100	336 h	32
Ac-TTAV-NH_2 4-mer/β-strand-directed library	M-AT	1 : 100	1 h	38
Ac-TTAI-NH_2[f] 4-mer/β-strand-directed library	M-AT Z-AT	1 : 100 1 : 10	1 h	38
Ac-TTAF-NH_2 4-mer/β-strand-directed library	M-AT	1 : 100	1 h	38
Ac-TTAY-NH_2 4-mer/β-strand-directed library	M-AT	1 : 100	1 h	38
Ac-TTAT-NH_2 4-mer/β-strand-directed library	M-AT	1 : 100	1 h	38
Ac-TTAL-NH_2 4-mer/β-strand-directed library	M-AT	1 : 100	1 h	38
Ac-TTAA-NH_2 4-mer/β-strand-directed library	M-AT	1 : 100	1 h	38
Ac-TTAM-NH_2 4-mer/β-strand-directed library	M-AT	1 : 100	1 h	38

[a] Amino acids are shown in one-letter code except norleucine (Nle). Abbreviations of N-terminal modifications: acetylated (Ac); formylated (formyl). [b] Abbreviations of serpins: α_1-antitrypsin (AT); antithrombin (ATIII); α_1-antichymotrypsin (ACT); plasminogen activator inhibitor-1 (PAI-1). [c] The molar ratios of serpin : peptide and incubation times were derived from different assays. Unless indicated elsewhere, all incubations were performed at 37 °C. [d] Ternary complexes were prepared by incubation of ATIII with Ac-SEAAAST-OH and WMDF, or Ac-SEAAAS-OH and Formyl-Nle-LF-OH, or with Ac-SEAAAS-OH and formyl-MLF peptides. [e] The ternary complex was formed by the glycosylated mutant of PAI-1 (Ala335Glu) with two molecules of the peptide Ac-TVASS-NH_2. [f] Interestingly, the identified peptide was found in the RCL sequence of protease nexin-1, but the peptide was not recognized by the serpin.

Notably, a 6-mer RCL peptide was used to explore the structural differences between the pathogenic Z-AT and normal M-AT.[29] The assessment of peptide binding was achieved by intrinsic tryptophan fluorescence and native PAGE as in previously related works. In addition, adding urea in native PAGE was able to distinguish the bound and unbound proteins unambiguously. Despite the sequence was derived from the RCL of AT, the 6-mer peptide preferentially annealed to Z-AT and did not significantly bind to other serpins (ATIII, ACT and PAI-1) that bear the similar tertiary structure. To further unravel how small peptides interact with serpins, around 40 RCL and exogenous peptides were selected for a separate study.[30] The acquired structural information represents a step forward in rational design of small molecules. However, no potent serpin-binding ligands have been identified from RCL derivatives, random selection and rational design to the best of our knowledge. A better screening approach is required to tackle this biological problem.

Large library from small libraries

Combinatorial library screening appeared to be an excellent alternative, provided sufficient structural information on the design of the library and an appropriate screening platform were available. A library screening method is normally tailor-made, and nondestructive analysis, *in situ* monitoring and no hindrance of protein–ligand interaction are some ideal pre-requisites. In addition, using all the 20 proteinogenic amino acids may be an aesthetic goal, but some other problems may arise and lead to futile screening such as the potential solubility issue. To this end, a series of small libraries derived from the aforementioned 6-mer peptide were synthesized to probe the

structural requirements and facilitated the design of a larger library.[32] The screening of the library was based on conventional polyacrylamide gel electrophoresis (PAGE). Initially, native-PAGE was used for the screening, however, unbound protein and the binary complex could not be distinguished unambiguously if the binding peptide has no charged amino acid. Native-PAGE in conjunction with heat-induced polymerization was able to identify the binding ligand, but the condition was too harsh and may disrupt the native state of protein and produce artefacts. Finally, urea was adulterated in native-PAGE and generated distinct band shifts. The effect of the chaotropic reagent was imperative and was used for the screening of the mixture-based library. A representative result of this conformation-sensitive assay is shown in Fig. 2.

The 6-mer peptide (Ac-FLEAIG-OH) corresponding to the P7-P2 sequence of the RCL was selected as the benchmark and starting point of the alanine scanning library. When compared with the RCL sequence of human serpins, P7-P2 is the only region that is not conserved and is likely to play a pivotal role in both binding affinity and specificity for each protease inhibitor system.[31] The smallest chiral amino acid alanine was used to substitute each non-alanine residue based on the 6-mer peptide one at a time to construct the focused library for screening. This library allowed the rapid determination of each individual residue's contribution to the binding and identified two extra 6-mer peptides (E3A: Ac-FLAAIG-OH and I5A: Ac-FLEAAG-OH) that bound to M and Z-AT.[32]

The two identified 6-mer peptides were used to develop their corresponding truncation libraries.[33] The incremental truncation of the two peptides (E3A and I5A) from both N- and C-terminal ends leads to the identification of the shortest peptide (Ac-FLAA-NH_2) that was able to bind to Z-AT and thus defined the minimum length.[31] The effect of N-terminal acetylation and C-terminal modifications was also assessed. The N-terminal acetylated peptide amide bound to AT more tightly than its corresponding N-terminal free and carboxyl-terminus peptides. The higher affinity is likely due to the extra hydrogen bond acceptor on the acetyl group and the elimination of the negative charge on the C-termini and thus resembles more an internal peptide segment. These findings were imperative, as fine-tuning binding affinity by manipulation of association and dissociation rates could be therapeutically advantageous.[34] The representative results (see ref. 32 for M-AT results) of alanine scanning and truncation libraries are shown in Fig. 3. The effect of D-amino acids was also evaluated and none of the D-substituted peptides were able to form a binary complex, significantly suggesting the presence of a stereochemical constraint binding interface. Through the systematic screening and stepwise evolution of focused libraries, the acetylated 4-mer peptide amide backbone was determined for the design of a large library.

The next step was the selection of amino acids for the mixture-based library. Insights into the inhibitory mechanism of serpins provided an excellent criterion for the selection of library building blocks. Upon snaring the cognate enzyme, the RCL of inhibitory serpin incorporates into the β-sheet and renders it a fully anti-parallel β-sheet conformation. Therefore, the residues within the RCL must adopt a conformation that is favored by the β-sheet to complete the inhibitory processes and are most likely the β-sheet preferred amino acids. To confirm this hypothesis, the RCL sequence of human serpins were analyzed in conjunction with the Chou–Fasman helix and sheet propensities.[35] The analysis showed that 72.8% of RCL residues (P7-P2) of inhibitory serpins consist of the top 10 β-sheet forming amino acids (A, H, I, L, M, F, T, V, W and Y).

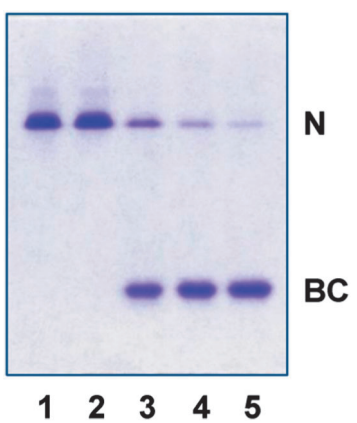


Fig. 2 Native-PAGE containing urea generating distinct band shifts between the binary complex (BC) and the peptide-free protein (N). α_1-Antitrypsin was incubated with and without (lane 1) a 100-fold molar excess of the peptide Ac-FLAA-NH_2 at 37 °C for 1 min (lane 2), 1 d (lane 3), 2 d (lane 4) and 3 d (lane 5). The more anodal migration of BC than N is due to the acquisition of extra conformation stability from the fully antiparallel β-sheet A.

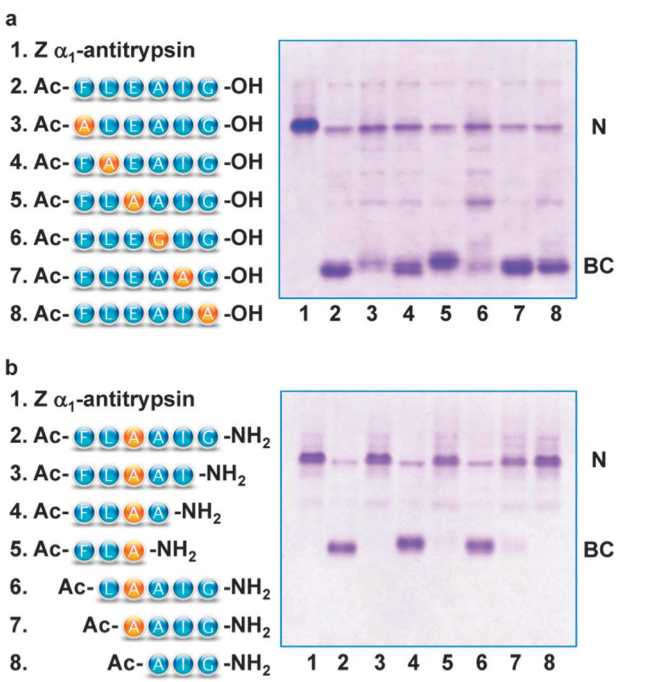


Fig. 3 Native-PAGE containing urea demonstrating the effect of (a) alanine scanning of the peptide Ac-FLEAIG-OH and its derived (b) truncation peptides on its binding to Z α_1-antitrypsin. The protein (N) was incubated with a 100-fold molar excess of the peptides at 37 °C for 3 d. Peptides capable of binding to the protein exhibit clear binary complexes (BC) and lesser ladder bands of the protein aggregates on the gel.

In contrast, only 38.3% of the residues were found in the RCL of non-inhibitory serpins. The diverse trend supports the hypothesis, and thus the 10 amino acids were selected for the construction of the mixture-based library.

Solid-phase split-and-mix synthesis was employed to synthesize the β-strand-directed library, and a simplified presentation of a library ($3 \times 3 \times 3$) is shown in Fig. 4.[36] In the case of the first generation of library (designated as Ac-$X_1X_2X_3X_4$-NH_2), 10 amino acids were used for the synthesis of the 4-mer peptide library and this represents a $10 \times 10 \times 10 \times 10$ library (10 000 peptides). The synthesis of the library was initiated at the X_4 position and terminated at the X_1 position. The scaffold and building blocks are shown in Fig. 5a. Upon completion of the 4th reaction cycle (coupling of the amino acids at the X_1 position), the generated 10 sub-libraries (each contains 1000 peptides) were not mixed and hence the N-terminal amino acids (X_1) were known. The 10 sub-libraries with diversity in the last 3 residues (X_2, X_3 and X_4) were represented as Ac-$AX_2X_3X_4$-NH_2, Ac-$FX_2X_3X_4$-NH_2, Ac-$HX_2X_3X_4$-NH_2, Ac-$IX_2X_3X_4$-NH_2, Ac-$LX_2X_3X_4$-NH_2, Ac-$MX_2X_3X_4$-NH_2, Ac-$TX_2X_3X_4$-NH_2, Ac-$VX_2X_3X_4$-NH_2, Ac-$WX_2X_3X_4$-NH_2 and Ac-$YX_2X_3X_4$-NH_2. Each randomized position (Xi) was a combination of the 10 amino acids (A, H, I, L, M, F, T, V, W and Y).

The results of library screening defined the next generation of library until all the X positions were revealed by multiple-step screening. Iterative deconvolution was employed to stepwise identify the optimal residues for both M and Z-AT binding.[37] Four cycles of syntheses and screenings were exhaustively repeated, and the sizes of libraries were decreased exponentially until all the peptide residues were defined in the final library (from 10 000 to 1000 and then 100 peptide mixtures, and finally 10 separate peptides). In other words, the first three rounds of screening were performed on the predefined mixtures, only the last screening was against 10 separate peptides and hence the optimal sequence could be identified. As shown in Fig. 5b, the first three cycles of screening identified the residues of T, T and A for the bindings to Z-AT from the libraries of Ac-$X_1X_2X_3X_4$-NH_2, Ac-$TX_2X_3X_4$-NH_2 and Ac-TTX_3X_4-NH_2, respectively. In the last cycle of screening, the optimal Z-AT-binding peptide was revealed as Ac-TTAI-NH_2 from the final library of Ac-$TTAX_4$-NH_2.[38]

The potent binding affinity and specificity of the identified Z-AT-binding peptide were demonstrated as it only required 10-fold molar excess of Ac-TTAI-NH_2 and 1 hour of incubation to form a clear binary complex (Fig. 5b; panel IV).[38] The optimal M-AT-binding peptide Ac-TTAF-NH_2 was also identified, which has the potential to develop a purification tag. Most recently, the binding of 55 peptides based on the sequence of strand 5 of AT was assessed to challenge the β-hairpin domain swapping model.[39] Only the peptide Ac-TTAI-NH_2 was able to bind to AT and thus supported the classical loop–sheet polymerization model. However, the peptide was mistakenly described as a RCL fragment, rather than a combinatorially selected peptide. A comparison of the binding affinity of RCL-derived, exogenous and combinatorially selected peptides is summarized in Table 1. The first RCL-derived peptide (14-mer) used to study the structural transition of AT required a high molar ratio

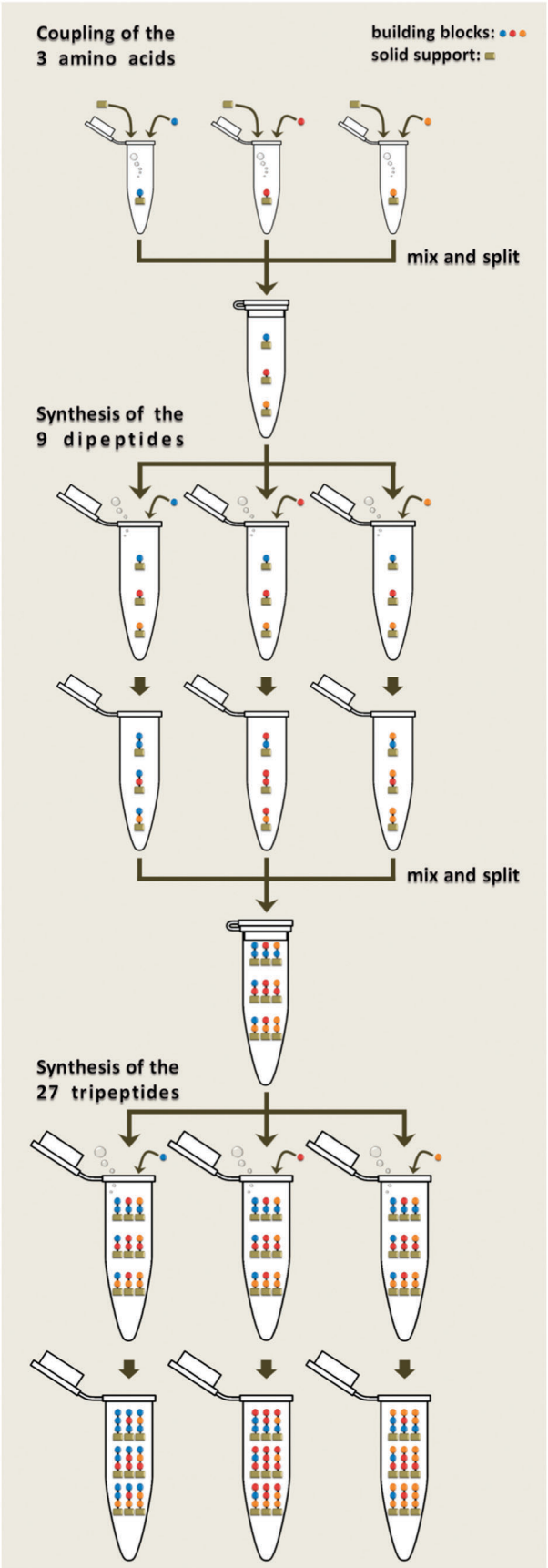


Fig. 4 The split-and-mix method showing a $3 \times 3 \times 3$ library (27 peptides) can be achieved in 3 synthetic steps. The size of the library could be increased exponentially by using more building blocks and synthetic cycles.

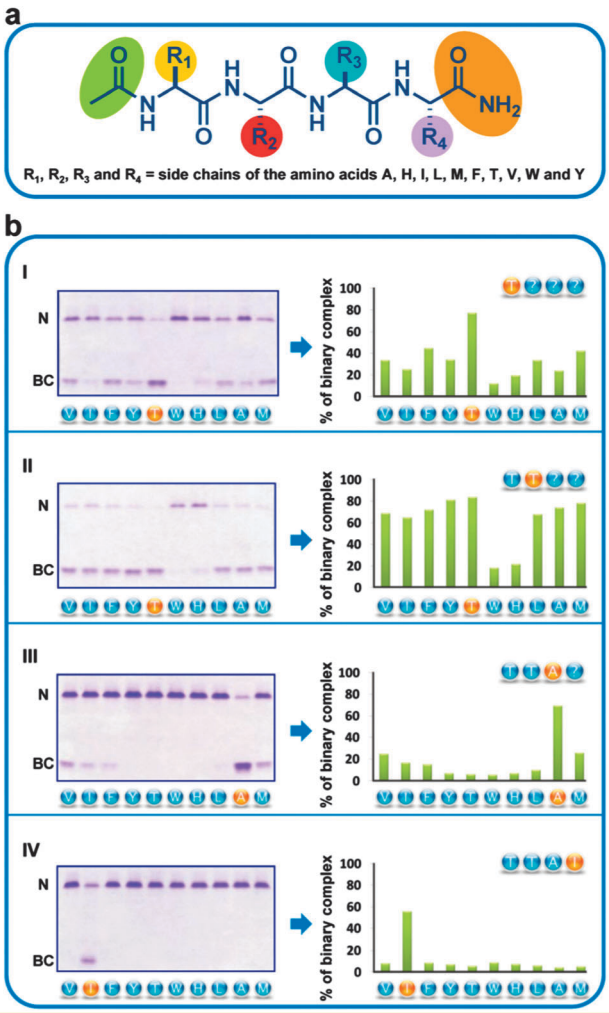


Fig. 5 (a) The scaffold and building blocks of the β-strand-directed library and (b) iterative deconvolution of Z α_1-antitrypsin (see ref. 32 for M-AT results). The potency of each sub-library was determined by the ability to form the binary complex (BC) as screened by native-PAGE containing urea, and further quantitatively measured to give the percentage of BC formation by densitometric analysis. The most reactive sub-library in each screening cycle is indicated on the top right corner. All screenings were performed with a calculated 10-fold molar excess of each individual peptide against Z α_1-antitrypsin at 37 °C for 2 h except the last cycle was only 1 h.

(normally 100 to 200-fold molar excess) and long incubation time (usually days) to bind to M-AT.[40–42] Some other serpin (AT, ATIII and PAI-1) RCL-derived peptides (8 to 14-mer) have also been synthesized to probe the inhibitory function, conformational stability and polymerization mechanism of serpins but exhibited similar binding affinity.[18,43–54] A shorter RCL-derived peptide (6-mer) was identified to preferentially anneal to Z-AT, facilitating the identification of other AT-binding peptides (4 to 6-mer).[29,32] Exogenous and other small RCL-derived peptides have also been selected to block and reverse serpin polymerization.[30] These RCL-derived peptides were valuable for functional and structural interrogation of serpins, but not suitable for the development of mimetics when considering the size and affinity. Moreover, promiscuous binding have been reported, such as ATIII RCL peptide binds to ATIII, ACT and AT.[29,30,50,52–54]

The combinatorially selected peptide Ac-TTAI-NH_2 bound far more tightly to Z-AT than those RCL-derived and exogenous peptides. The required molar ratio and incubation time to form BC were only 1 : 10 and 1 hour, respectively. In addition, the identified peptide preferentially annealed to Z-AT and was not recognized by M-AT and other serpin such as protease nexin-1 under the same conditions.[38]

The positional scanning library is a useful tool to screen amino acid(s) of interest at given position(s) and was used to validate the screening results of the β-strand-directed library.[55] A potential issue of the split-and-mix and iterative deconvolution methods is that the most potent compound may not be identified due to synthetic or screening artifacts. These effects should be taken into account particularly if the potency of each sub-library could not be distinguished unambiguously. Examining the screening results of the β-strand-directed library, the second residue (X_2) was not easily revealed without the assistance of densitometric analysis quantitatively.[38] Although threonine was eventually incorporated for the syntheses and identified the potent M and Z-AT binders, the binding affinity was suspected if it could be enhanced further by the other amino acids at the X_2 residue. Therefore, a positional scanning library (Ac-TX_2AF-NH_2; X_2 = A, H, I, L, M, F, T, V, W and Y) was prepared and screened against M-AT. This small focused library (10 separate peptides) was actually part of the β-strand-directed library Ac-T$X_2X_3X_4$-NH_2 (a mixture of 1000 peptides). These 10 peptides can be screened to examine if the identified residue at X_2 was due to a synergy of several peptides. The positional scanning-retrieved peptide, Ac-TTAF-NH_2, was identified again as the previous result of iterative deconvolution of M-AT (an unpublished result of Y.-P. Chang, and Y.-H. Chu). This result suggested that threonine was indeed the most potent residue and the selection at the X_2 position was appropriate.

Characterization of the combinatorally selected peptide

The combinatorially selected peptide Ac-TTAI-NH_2 was further assessed and explored by biophysical, cellular and computational methods. Surface plasmon resonance (SPR) technology was employed to evaluate the binding characteristics of the AT-binding peptide. In addition, this chip-based biosensor served as an excellent platform to validate the gel-based library screening orthogonally. The monitoring of protein–ligand interaction by SPR is real-time and without the need of tags or labels that might interfere with the binding interaction. We have previously used SPR to study the protein–peptide interactions of streptavidin and VanX between combinatorially selected cyclopeptides and dipeptide phosphonates, respectively.[56,57] In the validation of the Z-AT-binding peptide, the macromolecular Z-AT was deliberately immobilized onto the sensor chip to avoid any hindrance of protein–peptide interactions that might result from the immobilization chemistry. The protein chip is reusable and applicable to screening of peptides and small molecules if the regeneration protocol was optimized. The sensorgrams showed that the binding of the

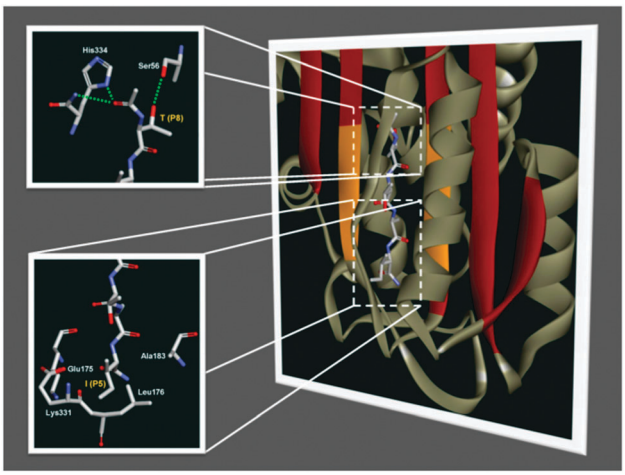


Fig. 6 Proposed structure of the complex of the peptide Ac-TTAI-NH_2 and AT. The NH and CO groups from the backbone of the incorporated peptide are hydrogen bonded to the backbones of adjacent s3A and s5A (marked in orange) of α_1-antitrypsin and render the β-sheet into a 6-stranded antiparallel β-sheet. The inset on the upper left shows the hydrogen bond (light green dashed line) between the N-terminal Thr (P8) of peptide Ac-TTAI-NH_2 and Ser56. Additional hydrogen bonds derived from the acetyl group of the peptide with the side chain of His334 (light green dashed line) and the backbone NH group of Lys335 (light green dashed line) are also illustrated. The inset on the lower left shows the hydrophobic side chain of Ile interacts with a pocket surrounded by the residues of Val173, Glu175, Leu176, Ala183 and Lys331. Carbon, nitrogen and oxygen atoms are shown in white, light blue and red, respectively.

identified peptide was specific to Z-AT and in a dose-dependent manner. In addition, the slow association corresponds to the incubation time required to form BC, while the slow dissociation rate contributes to the tight binding and correlates with the gel-based assay.[38]

The cytotoxicity of the identified peptide was determined by MTT assay in two normal lung epithelial cell lines (BEAS-2B and NL20), two normal lung fibroblast cell lines (WI-38 and IMR-90) and one cancer cell line (A2058). The viability of cells was not affected even at high concentrations, which indicates that the small peptide is not cytotoxic.[38] Most recently, it has been shown that the peptide not only blocks Z-AT polymerization, but also reverses the pathogenic aggregation *in vivo*.[58] In addition, the induction of PERK (protein kinase RNA-like ER kinase)-dependent NF-jB, IL-6, IL-8, and RGS16 and calnexin were abrogated effectively by the peptide, which means the ER stress response due to Z-AT aggregation was also alleviated.

The structure of the binary complex was proposed to elucidate the binding interaction between the identified peptide and the s4A site of AT.[38] In brief, the peptide Ac-TTAI-NH_2 inserts into the lower part of the β-sheet A and lines up with the residues on strands 3 (Ala183, Leu184, Val185 and Asn186) and 5 (Lys331, Ala332, Val333 and His334) as the central strand 4. As shown in Fig. 6, additional hydrogen bonds are found at the N-terminal threonine and its acetyl group with the side chain of His334, the backbone NH of Lys335 and the side chain of Ser56, respectively. The N-terminal threonine may act as the ''anchoring residue'' to locate the peptide and lands at the hot spot. Subsequently, the flexible peptide chain may pinpoint and bind to the β-sheet A through those backbone hydrogen bonds. According to the Venn diagram of amino acids, all the residues are in the category of small amino acids up to the penultimate residue of the peptide.[59] These smaller side chains likely facilitate the incorporation of the β-sheet A by evading the connecting loop of the F-helix, which was reported to undergo a reversible conformational change during the inhibitory mechanism.[60] Finally, the bulky side chain of the C-terminal isoleucine may occupy the cavity surrounded by Val173, Leu176, Ala183, Glu175 and Lys331. Taken together, the tight binding of the binary complex is stabilized by hydrogen bonds, hydrophobic interactions and cavity-filling effect.

Conclusions

To date, there is no cure or amiable treatment for AATD and other conformational disease patients, only supplemental treatments temporarily alleviate the symptoms. The predicament of conformational disease underscores the need to develop effective treatment. In the past few years, several promising strategies have been suggested for AATD, such as gene therapy, promotion of hepatic AT secretion, inhibition of NE and inhibition of protein polymerization. A recent biological study showed that induced pluripotent stem cells were able to correct the Z mutation in mouse.[61] The chemical approach described herein represents a systematic and flexible screening platform and has facilitated the identification of potent anti-protein aggregation inhibitors. The effectiveness of the ''libraries from libraries'' concept was demonstrated by the evolution of the initial alanine scanning library to the truncation, D-amino acid scanning and eventually the β-strand-directed library and its related positional scanning library. These libraries are complementary to each other and identified the optimal peptide sequences accordingly, which are useful for future mimetic design. The conformation-sensitive assay, native-PAGE containing urea was used throughout the screening and the effect of the chaotropic reagent immensely improved the readout of gels. Characterization of the combinatorially selected inhibitor was achieved by biophysical, cellular and computational methods. These data demonstrate the potential to ameliorate both liver and pulmonary dysfunctions resulting from abnormal protein aggregation. Taken together, the developed chemical approach paves the way for the discovery of anti-protein aggregation inhibitors, which is conceivably feasible for other conformational diseases.

Acknowledgements

We thank Chien-Hung Chen, Kai-Chun Lin and Sheng-Chieh Lin from laboratories of Dr Wun-Shaing Wayne Chang (National Health Research Institutes, Taiwan) for their assistance in protein purification and cellular work. We also thank for the circular dichroism data from the research group of Dr Timothy R. Dafforn (University of Birmingham, UK). We are particularly grateful to Dr Ravi Mahadeva (University of Cambridge, UK) and Dr Wun-Shaing Wayne Chang (National Health Research Institutes, Taiwan) for helpful discussions throughout the course of the study. This work was supported

by grants from the National Science Council of Taiwan, ROC (NSC100-2113-M-194-003-MY3 and NSC97-2113-M-194-006-MY3) and an excellent research award from National Chung Cheng University (CCU101-00885-01). We thank reviewers for constructive comments and valuable suggestions.

Notes and references

1 F. Chiti and C. M. Dobson, *Annu. Rev. Biochem.*, 2006, **75**, 333.
2 J. Xu, J. Reumers, J. R. Couceiro, F. D. Smet, R. Gallardo, S. Rudyak, A. Cornelis, J. Rozenski, A. Zwolinska, J.-C. Marine, D. Lambrechts, Y.-A. Suh, F. Rousseau and J. Schymkowitz1, *Nat. Chem. Biol.*, 2011, **7**, 285.
3 B. Gooptu and D. A. Lomas, *Annu. Rev. Biochem.*, 2009, **78**, 147.
4 J. K. Stoller and L. S. Aboussouan, *Arch. Intern. Med.*, 2009, **169**, 546.
5 J. K. Stoller and L. S. Aboussouan, *Lancet*, 2005, **365**, 2225.
6 C. B. Laurell and S. Eriksson, *Scand. J. Clin. Lab. Invest.*, 1963, **15**, 132.
7 S. Eriksson, *Acta Med. Scand.*, 1964, **175**, 197.
8 A. K. Kubba and M. Young, *Chest*, 1998, **113**, 210.
9 V. Boulyjenkov, *Bull. W. H. O.*, 1997, **75**, 397.
10 E. D. Werner, J. L. Brodsky and A. A. McCracken, *Proc. Natl. Acad. Sci. U. S. A.*, 1996, **93**, 13797.
11 R. R. Kopito and D. Ron, *Nat. Cell Biol.*, 2000, **2**, E207.
12 R. W. Carrell, J. O. Jeppsson, C. B. Laurell, S. O. Brennan, M. C. Owen, L. Vaughan and D. R. Boswell, *Nature*, 1982, **298**, 329.
13 A. M. Wood and R. A. Stockley, *Respiration*, 2007, **74**, 481.
14 F. J. de Serres, *Chest*, 2002, **122**, 1818.
15 J. Whisstock, R. Skinner and A. M. Lesk, *Trends Biochem. Sci.*, 1998, **23**, 63.
16 J. A. Huntington, R. J. Read and R. W. Carrell, *Nature*, 2000, **407**, 923.
17 P. Sivasothy, T. R. Dafforn, P. G. W. Gettins and D. A. Lomas, *J. Biol. Chem.*, 2000, **275**, 33663.
18 D. A. Lomas, D. L. Evans, J. T. Finch and R. W. Carrell, *Nature*, 1992, **357**, 605.
19 B. Gooptu, B. Hazes, W.-S. W. Chang, T. R. Dafforn, R. W. Carrell, R. J. Read and D. A. Lomas, *Proc. Natl. Acad. Sci. U. S. A.*, 2000, **97**, 67.
20 D. Bruce, D. J. Perry, J.-Y. Borg, R. W. Carrell and M. R. Wardell, *J. Clin. Invest.*, 1994, **94**, 2265.
21 K. S. Aulak, E. Eldering, C. E. Hack, Y. P. T. Lubbers, R. A. Harrison, A. Mast, M. Cicardi and A. E. Davis III, *J. Biol. Chem.*, 1993, **268**, 18088.
22 R. L. Davis, A. E. Shrimpton, P. D. Holohan, C. Bradshaw, D. Feiglin, G. H. Collins, P. Sonderegger, J. Kinter, L. M. Becker, F. Lacbawan, D. Krasnewich, M. Muenke, D. A. Lawrence, M. S. Yerby, C. M. Shaw, B. Gooptu, P. R. Elliott, J. T. Finch, R. W. Carrell and D. A. Lomas, *Nature*, 1999, **401**, 376.
23 P. E. Stein, A. G. W. Leslie, J. T. Finch, W. G. Turnell, P. J. McLaughlin and R. W. Carrell, *Nature*, 1990, **347**, 99.
24 J. Mottonen, A. Strand, J. Symersky, R. M. Sweet, D. E. Danley, K. F. Geoghegan, R. D. Gerard and E. J. Goldsmith, *Nature*, 1992, **355**, 270.
25 R. W. Carrell, P. E. Stein, G. Fermi and M. R. Wardell, *Structure*, 1994, **2**, 257.
26 A. Wei, H. Rubin and B. S. Cooperman, *Nat. Struct. Biol.*, 1994, **1**, 251.
27 P. R. Elliott, D. A. Lomas, R. W. Carrell and J. P. Abrahams, *Nat. Struct. Biol.*, 1996, **3**, 676.
28 Y.-P. Chang, R. Mahadeva, A. O. M. Patschull, I. Nobeli, U. I. Ekeowa, A. R. McKay, K. Thalassinos, J. A. Irving, I. U. Haque, M. P. Nyon, J. Christodoulou, A. Gonzalez, E. Miranda and B. Gooptu, *Methods Enzymol.*, 2011, **501**, 139.
29 R. Mahadeva, T. R. Dafforn, R. W. Carrell and D. A. Lomas, *J. Biol. Chem.*, 2002, **277**, 6771.
30 A. Zhou, P. E. Stein, J. A. Huntington, P. Sivasothy, R. W. Carrell and D. A. Lomas, *J. Mol. Biol.*, 2004, **342**, 931.
31 P. G. W. Gettins, *Chem. Rev.*, 2002, **102**, 4751.
32 Y.-P. Chang, R. Mahadeva, W.-S. W. Chang, A. Shukla, T. Dafforn and Y.-H. Chu, *Am. J. Respir. Cell Mol. Biol.*, 2006, **35**, 540.
33 M. Ostermeier, A. E. Nixon, J. H. Shim and S. J. Benkovic, *Proc. Natl. Acad. Sci. U. S. A.*, 1999, **96**, 3562.
34 H. Parfrey, T. R. Dafforn, D. Belorgey, D. A. Lomas and R. Mahadeva, *Am. J. Respir. Cell Mol. Biol.*, 2004, **31**, 133.
35 P. Y. Chou and G. D. Fasman, *Annu. Rev. Biochem.*, 1978, **47**, 251.
36 Á. Furka, *Drug Discovery Today*, 2002, **7**, 1.
37 R. A. Houghten, C. Pinilla, S. E. Blondelle, J. R. Appel, C. T. Dooley and J. H. Cuervo, *Nature*, 1991, **354**, 84.
38 Y.-P. Chang, R. Mahadeva, W.-S. W. Chang, S.-C. Lin and Y.-H. Chu, *J. Cell. Mol. Med.*, 2009, **13**, 2304.
39 U. I. Ekeowa, J. Freekeb, E. Miranda, B. Gooptu, M. F. Bush, J. Pérez, J. Teckman, C. V. Robinson and D. A. Lomas, *Proc. Natl. Acad. Sci. U. S. A.*, 2010, **107**, 17146.
40 A. J. Schulze, U. Baumann, S. Knof, E. Jaeger, R. Huber and C.-B. Laurell, *Eur. J. Biochem.*, 1990, **194**, 51.
41 A. J. Schulze, R. Huber, E. Degryse, D. Speck and R. Bischoff, *Eur. J. Biochem.*, 1991, **202**, 1147.
42 A. J. Schulze, P. W. Frohnert, R. A. Engh and R. Huber, *Biochemistry*, 1992, **31**, 7560.
43 I. Björk, K. Ylinenjärvi, S. T. Olson and P. E. Bock, *J. Biol. Chem.*, 1992, **267**, 1976.
44 R. W. Carrell, D. L. Evans and P. E. Stein, *Nature*, 1991, **353**, 576.
45 D. A. Lomas, D. L. Evans, S. R. Stone, W.-S. W. Chang and R. W. Carrell, *Biochemistry*, 1993, **32**, 500.
46 H. L. Fitton, R. N. Pike, R. W. Carrell and W.-S. W. Chang, *Biol. Chem.*, 1997, **378**, 1059.
47 D. C. Crowther, L. C. Serpell, T. R. Dafforn, B. Gooptu and D. A. Lomas, *Biochemistry*, 1992, **31**, 2720.
48 D. T. Eitzman, W. P. Fay, D. A. Lawrence, A. M. Francis-Chmura, J. D. Shore, S. T. Olson and D. Ginsbur, *J. Clin. Invest.*, 1995, **95**, 2416.
49 J.-O. Kvassman, D. A. Lawrence and J. D. Shore, *J. Biol. Chem.*, 1995, **270**, 27942.
50 W.-S. W. Chang, M. R. Wardell, D. A. Lomas and R. W. Carrell, *Biochem. J.*, 1996, **314**, 647.
51 Y. Xue, P. Björquist, T. Inghardt, M. Linschoten, D. Musil, L. Sjölin and J. Deinum, *Structure*, 1998, **6**, 627.
52 R. Skinner, W.-S. W. Chang, L. Jin, X. Pei, J. A. Huntington, J. P. Abrahams, R. W. Carrell and D. A. Lomas, *J. Mol. Biol.*, 1998, **283**, 9.
53 D. C. Crowther, L. C. Serpell, T. R. Dafforn, B. Gooptu and D. A. Lomas, *Biochemistry*, 2003, **42**, 2355.
54 A. Zhou, P. E. Stein, J. A. Huntington and R. W. Carrell, *J. Biol. Chem.*, 2003, **278**, 15116.
55 C. Pinilla, J. R. Appel, P. Blanc and R. A. Houghten, *BioTechniques*, 1992, **13**, 901.
56 Y.-P. Chang and Y.-H. Chu, *Anal. Biochem.*, 2005, **340**, 74.
57 Y.-P. Chang, M.-J. Tseng and Y.-H. Chu, *Anal. Biochem.*, 2006, **359**, 63.
58 S. Alam, J. Wang, S. Janciauskiene and R. Mahadeva, *J. Hepatol.*, 2012, **57**, 116.
59 W. R. Taylor, *J. Theor. Biol.*, 1986, **119**, 205.
60 L. D. Cabrita, W. Dai and S. P. Bottomley, *Biochemistry*, 2004, **43**, 9834.
61 K. Yusa, S. T. Rashid, H. Strick-Marchand, I. Varela, P.-Q. Liu, D. E. Paschon, E. Miranda, A. Ordóñez, N. R. F. Hannan, F. J. Rouhani, S. Darche, G. Alexander, S. J. Marciniak, N. Fusaki, M. Hasegawa, M. C. Holmes, J. P. Di Santo, D. A. Lomas, A. Bradley and L. Vallier, *Nature*, 2011, **478**, 391.

