Protein folding and calcium binding defects arising from familial hypercholesterolemia mutations of the LDL receptor

Studies of a critical module in the ligand binding domain of the LDL receptor implicate a protein folding defect, coupled to a deficiency in calcium binding, as a major cause of familial hypercholesterolemia.

Sir—The low-density lipoprotein receptor (LDLR) is a large, membrane-bound, modular protein responsible for the uptake of cholesterol-containing lipoprotein particles into cells[1]. The N-terminal domain of the LDLR, responsible for ligand binding, consists of seven tandemly repeated modules[2,3], each ~40 residues long and containing three disulphide bonds (Fig. 1). These 40-residue repeats are also found in numerous other proteins, including the LDL receptor-related protein, components of the complement cascade (C6–C9), and the cell-surface receptor for Rous sarcoma virus[4–9].

Over 150 mutant alleles[10] of the low density lipoprotein receptor (LDLR) gene have been identified in patients with familial hypercholesterolemia (FH)[2,3]. Many of these alleles are point mutations scattered among the seven ligand-binding repeats of the LDLR, with more clustered in repeat 5 (LR5) than elsewhere (see Fig. 1). Here, we show that isolated LR5 folds efficiently *in vitro* to a single disulphide-bonded isomer in the presence of calcium. In contrast, variants of repeat 5 that contain point mutations identified in FH patients fail to form a unique disulphide-bonded species, implicating a protein folding defect as a major cause of FH. Consideration of the locations and the effects of the point mutations suggests that the conserved DCxDxSDE acidic sequence of each repeat may not direct interaction with a positively charged surface of the ligand, as previously suggested[1,2,11], but may correspond instead to a calcium-binding motif that is crucial for protein folding.

Calcium dependent folding of LR5

Isolated repeat 5 (LR5) of the human LDLR was produced in bacteria and purified by HPLC. In the presence of calcium under conditions that permit disulphide exchange, wild-type LR5 folds to a single isomer (Fig. 2*b*). In the absence of calcium, a distribution of disulphide-bonded isomers is formed (Fig. 2*a*), resembling that obtained in denaturant (6M guanidine HCl; data not shown). LR5 oxidized under native conditions has a disulphide connectivity of C1–C3, C2–C5, and C4–C6, consistent with that reported for ligand-binding repeats 1 (ref. 12) and 2 (ref. 13). The calcium affinity of LR5 with native disulphide bonds, as measured by fluorescence, is ~70 nM (Fig. 2*c*). Proton NMR spectra of LR5, with native disulphide bonds, were acquired before (Fig. 3*a*) and after (Fig. 3*b*) the addition of calcium. Only the spectrum taken after addition of calcium is characteristic of a folded protein, with significant chemical shift dispersion away from random-coil values in the amide region.

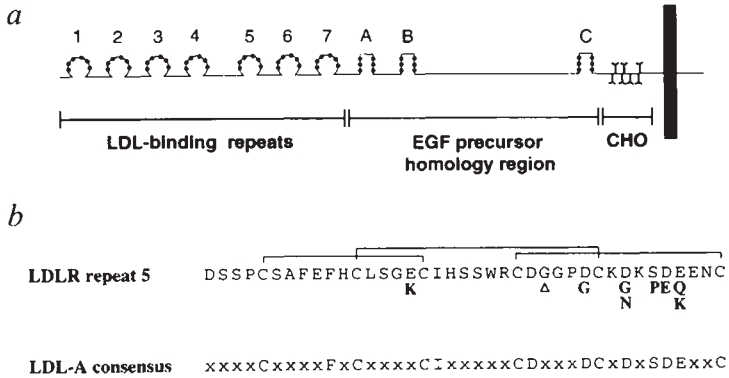


Fig. 1 *a*, Schematic diagram of the modular arrangement of the low-density lipoprotein receptor (LDLR). Ligand binding repeats (1–7), epidermal growth factor-like repeats (A–C), and glycosylation sites (CHO) are indicated. *b*, Sequence and disulphide connectivity of ligand-binding repeat 5 (LR5) of the LDLR. Point mutations, evaluated in this study, that give rise to the genetic disease familial hypercholesterolemia are indicated below the sequence in bold type. Consensus residues, based on sequence alignment of the seven ligand-binding repeats of the LDLR[2], are listed on the bottom row.

Folding defect in FH

Nine point mutants of LR5, each known to cause FH when in the intact receptor[10], were examined for the ability to fold to the native (wt) disulphide-bonded form. Under redox conditions that permit quantitative folding of wild-type LR5, none of these mutant LR5 proteins fold to a single disulphide isomer (Fig. 4*a*). Except for the D203G mutant, the distribution of isomers at equilibrium is similar when refolding is attempted in a redox buffer containing calcium (1 mM), lacking calcium (1 mM EDTA), or in denaturant (Fig. 4*a*). Moreover, the major species

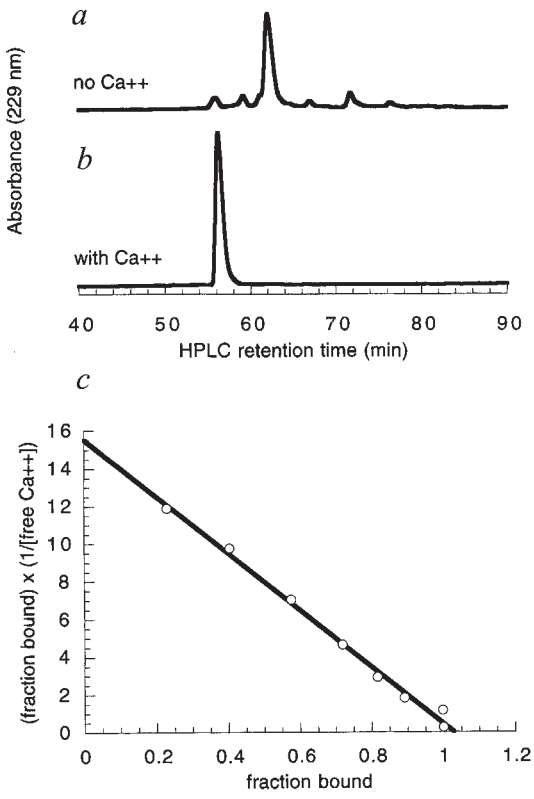


Fig. 2 Folding of ligand-binding repeat 5 (LR5) of the low-density lipoprotein receptor *in vitro* to the native disulphide-bonded isomer is calcium-dependent. *a*, HPLC chromatogram of products formed after folding, when calcium is not present, under conditions permitting disulphide exchange. A distribution of non-native disulphide-bonded species is formed. When LR5 is refolded in denaturant (6 M guanidine HCl), a very similar chromatogram is seen. *b*, HPLC chromatogram of products formed after folding in the presence of calcium, under conditions otherwise identical to those in (*a*). Of 15 possibilities, only the native disulphide isomer is formed. *c*, Scatchard analysis of calcium binding by LR5 of the LDL receptor. The best fit line (see Methods) gives an approximate K_d of 70 nM, and a stoichiometry of 1.03 calcium ions per LR5.

formed in each case contain non-native disulphide bonds (see Methods).

Our observation that FH-causing mutations of LR5 give rise to a protein-folding defect likely explains the delayed transport to the Golgi observed in previous pulse-chase studies of the intact receptor for three of the mutations studied here: deletion of residue G197 (ref. 14), and the D206E and E207K substitutions[15,16]. Our findings are also consistent with previous work evaluating the structural and functional consequences of the four residue deletion within the conserved DxSDE sequence (Fig. 1) in the Watanabe heritable hyperlipidemic (WHHL) rabbit[17]. The protein harbouring this deletion fails to be transported to the cell surface, but is recognized by conformation-specific antibodies directed to other regions of the protein and is capable of binding LDL, suggesting that the WHHL mutation gives rise to a folding defect localized to the cysteine-rich repeat in which it lies[18]. Protein folding defects have been implicated in several other human diseases[19], including cystic fibrosis[20], α-1-antitrypsin deficiency[21,22], retinitis pigmentosa[23], and Marfan's syndrome[24].

To determine whether an increased calcium concentration may overcome the folding defect for some of the LR5 variants, we examined the calcium dependence of folding for wild-type and mutant LR5 variants. At the lowest concentration of calcium examined (0.01 mM), greater than 50% of wild-type LR5 molecules fold to the native disulphide isomer (Fig. 4*b*). None of the mutants studied, other than D203G, form a significant proportion of the native species at calcium concentrations up to 10 mM.

For the D203G mutant, the concentration of calcium at which 50% of the molecules contain native disulphide bonds is 20-fold higher than for wild-type LR5 (Fig. 4*b*). The calcium dependence of native disulphide bond formation in this mutant bears a striking resemblance to a mutation from a patient with Marfan's syndrome (R1137P). This mutation, that lies within a calcium-binding EGF-like domain of fibrillin, also interferes with the folding of this domain by lowering the affinity of this EGF-like repeat for calcium[24].

A crucial calcium-binding motif

Remarkably, six of the nine folding-defective point mutations of LR5 alter acidic residues within a DCxDxSDE sequence (Fig. 1*b*) that is conserved near the C terminus of each ligand-binding repeat sequence. The seventh mutation, S205P, involves the conserved serine residue within this sequence. It is highly likely, therefore, that the folding defect we observe in these mutations results from the disruption of calcium binding, which is coupled to proper folding of LR5 by a 'calcium box' within the conserved DCxDxSDE sequence. How the E187K and the ΔG197 mutations result in improper folding of LR5 is not immediately evident.

The DCxDxSDE sequence is reminiscent of acidic clusters that characterize other calcium binding motifs, such as the EF-hand[25,26] ($DxDxDx_6E$) and the subset of EGF modules that coordinate calcium[27,28] (DxDE). Although it has been suggested previously that the conserved DCxDxSDE sequence might be critical for binding to

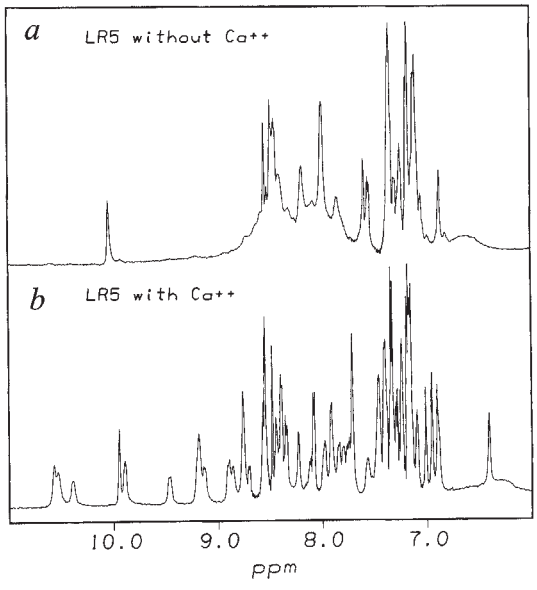
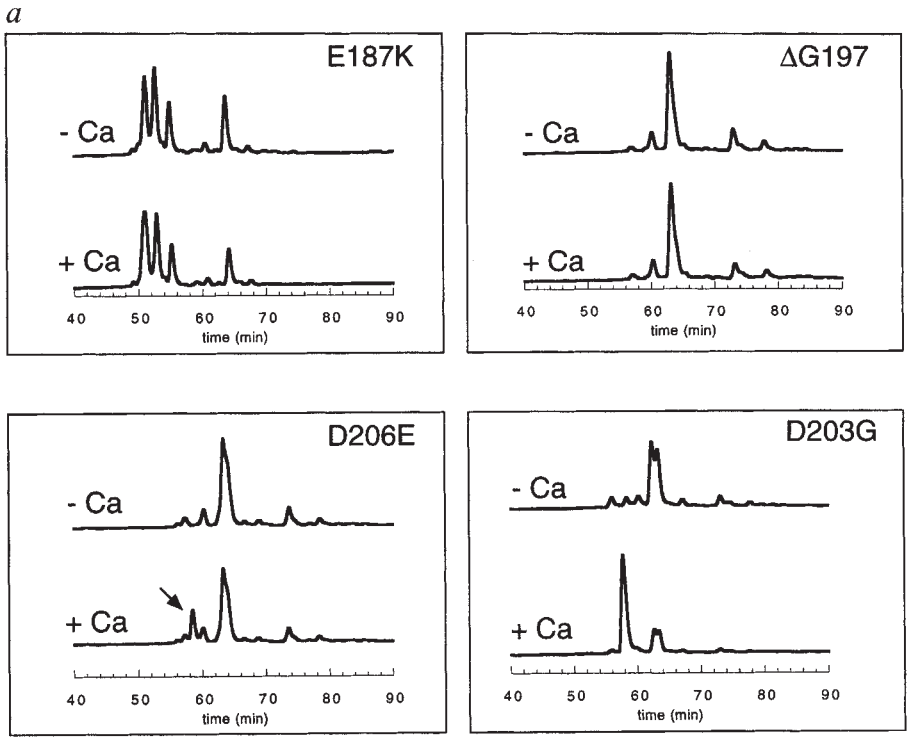


Fig. 3 *a*, ^{1}H NMR spectrum of ligand-binding repeat 5 of the low-density lipoprotein receptor in the absence of calcium at pH 5.0. The chemical shift dispersion is poor, and shifts do not deviate significantly from random-coil values[39]. The amide proton signals are much reduced in intensity in the spectrum taken in the absence of calcium, perhaps because presaturation of the water causes some of the exchangeable amides in the unfolded protein to disappear. *b*, ^{1}H NMR spectrum of LR5 after the addition of 10 mM calcium at pH 5.0. The enhanced dispersion of amide chemical shifts (6.5–10.5 p.p.m.) sharply contrasts with the spectrum in (*a*), and is indicative of folded structure.



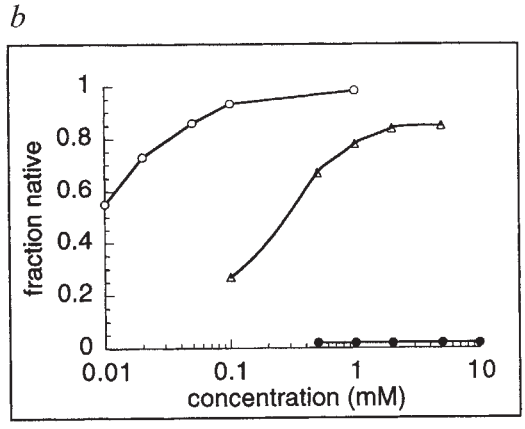


Fig. 4 Improper folding of point mutants of ligand-binding repeat 5 (LR5) of the low-density lipoprotein receptor. Wild-type LR5 refolds to a single disulphide isomer in the presence of calcium (Fig. 2*b*), and a different distribution of disulphide isomers is seen without calcium (Fig. 2*a*). *a*, Point mutants of LR5 (E187K, ΔG197, D206E and D203G) do not oxidize to a single disulphide isomer, even in the presence of calcium. Except for the D203G variant, the distribution of disulphide isomers is similar in the absence (top) or presence (bottom) of calcium (1 mM). Chromatograms similar to those of the ΔG197 mutant are obtained for the D200G, D203N, S205P, E207Q, and E207K LR5 point mutants (not shown). *b*, Fraction of molecules that fold to the native disulphide connectivity as a function of calcium concentration (0.01–10 mM). Wild-type LR5 (open circles), D203G (open triangles) and S205P (filled circles).

lipoprotein ligands by interacting with a positively charged patch on the fourth helix[29] of apolipoprotein E[1,2,11], our observations are more consistent with a structural role for this conserved sequence motif in the coordination of calcium. Our proposal is further supported by the NMR structures of both ligand-binding repeats 1 (ref. 30) and 2 (ref. 31). Inspection of these NMR structures indicates that the side chains of residues analogous to D200, D206, and E207 (DCxDxSDE) are positioned near enough to one another to act as ligands coordinating a single calcium ion. In contrast, the residue analogous to D203 lies on a surface loop distant from this potential calcium-binding site and would be unlikely to coordinate calcium directly, consistent with our observation that the D203G mutant retains the ability to form native disulphide bonds in the presence of higher concentrations of calcium.

In vivo correlation

Finally, our *in vitro* folding data with isolated LR5 correlate well with *in vivo* phenotype severity observed in cells expressing the full-length LDLR. Although no *in vivo* activity data have been reported for homozygotes harbouring the D203G mutation, fibroblasts from patients homozygous for the ΔG197, D203N, S205P, and E207K mutations, which fail to form detectable amounts of the native disulphide isomer *in vitro* (Fig. 4*a*), have undetectable LDL uptake activity[10]. Cultured fibroblasts from patients homozygous for the D206E mutation, which results in a small but significant fraction of native disulphide bonds (Fig. 4*a*, arrow), retain small amounts (5–15%) of LDL uptake activity[10].

Methods

Expression and purification of LR5. Plasmid pLDLR2[3], which contains the gene encoding the human LDLR, was obtained from the American Type Culture Collection (Rockville, MD). A cassette encoding residues 172–211 of the LDLR, flanked by Hind III and Bam H1 restriction sites, was constructed by the polymerase chain reaction using appropriate oligonucleotides, and was subcloned into the vector pMMHb (a kind gift of M. Milhollen). In the resulting plasmid, pMM-LR5, LR5 is expressed as a fusion with a modified form of the trpLE sequence in which the Met and Cys residues have been replaced by Leu and Ala, respectively[32], and an N-terminal $(His)_9$ tag has been added, creating a $(His)_9$-TrpLE-Met-LR5 chimeric protein. Oligonucleotide-directed mutagenesis[33] was used to create the variants of LR5 studied here. Standard recombinant DNA techniques were used throughout[34].

Inclusion bodies containing the chimeric proteins, expressed in *Escherichia coli* strain BL21(DE3) pLys(S)[35], were isolated and purified from the cell pellet as described[36], and then dissolved in 10 mM tris buffer, pH 8.0, containing 6.0 M guanidine HCl and 10 mM oxidized 2-mercaptoethanol. After oxidation overnight at 4 °C, the sample was loaded onto a column of Ni^{+2}-NTA agarose (Qiagen; Chatsworth, CA). The fusion protein was eluted with a solution of 0.2% acetic acid containing 6M guanidine HCl, dialysed exhaustively against 5% acetic acid, and lyophilized.

The lyophilized fusion protein was cleaved with cyanogen bromide, dialysed exhaustively against 5% aqueous acetic acid after cleavage, and lyophilized[32]. Purification to homogeneity by chromatography on DEAE sepharose (Pharmacia), followed by reversed-phase HPLC, was performed

as described[37]. The identities of wild-type and LR5 mutant peptides were confirmed by MALDI-TOF mass spectrometry (Voyager Elite, PerSeptive Biosystems). All observed masses were within 3 a.m.u. of the expected value. Purified peptides were stored in a dessicator at 4 °C.

Folding of wt and mutant LR5 variants under conditions permitting disulphide exchange. Reduced wild-type or mutant LR5 (10 μM) was refolded at room temperature in an anaerobic chamber (Coy Laboratory Products). Equilibration under native conditions was performed in 50 mM tris buffer, pH 8.5, containing reduced glutathione (500 μM), oxidized glutathione (250 μM), and calcium chloride (1 mM). Calcium-free equilibration was performed in the identical redox buffer, containing EDTA (1 mM), and equilibration of disulphide isomers under denaturing conditions was performed in the same buffer, containing 6 M guanidine HCl.

Aliquots of the refolding reactions were analysed by reversed-phase HPLC on a Vydac C-18 column using a linear water-acetonitrile gradient (0.1% per minute) containing 0.1% TFA. Comparison of 24 and 48 h time points suggests that an equilibrium distribution of disulphide-bonded isomers is reached after 24 h.

Disulphide connectivity of native LR5. The disulphide bond connectivity of the native disulphide isomer of wild-type LR5 was established by analysis of proteinase digestion fragments. Wt LR5 (~1 mg) was digested with proteinase K (10% w/w) in 1 ml of 50 mM tris buffer, pH 8.0, at 65 °C for 16 h, followed by analytical reversed-phase HPLC. Peptide fragments with altered elution times upon reduction with DTT (50 mM) were purified and identified by mass spectrometry (Voyager Elite, PerSeptive Biosystems) and N-terminal sequencing. The following peptide fragments were identified for wild-type LR5: (i) Asp 172–Cys 176/ Ser 185–His 190 (M_r measured 1150, expected 1151, corresponding to the C1–C3 disulphide bond) (ii) Ser 205–Ala 211/ Trp 193–Asp 196 (M_r measured 1346, expected 1344, corresponding to the C4–C6 disulphide bond). Digestion of ~50 μg wild-type LR5 at 37 °C for 8 h with Endo-Glu (Boeringer-Mannheim, 2 μg) in 20 μl sodium phosphate buffer (50 mM, pH 7.8) produces fragments of molecular weight 2071 and 1942 (predicted M_r values for the fragments Phe 181–Glu 187/Gly 197–Glu 208 and Phe 181–Glu 187/Gly 197–Glu 207 are 2069 and 1940 respectively), which disappear following reduction with DTT, indicating the presence of the C2–C5 disulphide bond. This assignment of the LR5 disulphide connectivity is consistent with those reported for ligand-binding repeats 1 and 2[12,13].

Disulphide connectivities of LR5 mutants. Reduced wt or mutant LR5 (10 μM) was allowed to fold at room temperature as in Fig. 2. Non-native disulphide isomers of wt and mutant LR5 proteins, sufficient for proteolysis with Endo-Glu (~50 μg), were purified after folding under these same conditions. For a typical digestion, ~50 μg purified lyophilized protein (native and non-native disulphide isomers of wt and mutant LR5) was incubated at 37 °C for 8 h in 20 μl of 50 mM sodium phosphate buffer, pH 7.8, containing 2 μg Endo-Glu. Aliquots were removed at 4 and 8 h, diluted 3-fold with water, and subjected to MALDI-TOF mass spectrometry (Voyager Elite, PerSeptive Biosystems). Digestion of wt LR5 yields two diagnostic fragments, of mass 2353 (expected mass 2351 for a fragment containing the C1–C3, and C4–C6 disulphides) and mass 2069 (expected mass 2071 for a fragment containing the C2–C5 native disulphide bond). In contrast, digestion of the predominant non-native disulphide-bonded isomer formed by LR5 after equilibration in 6 M guanidine HCl yields a digestion product of mass 3157 (expected mass 3156 for the Phe 181–Glu 208 fragment of LR5, which contains cysteines 2, 3, 4 and 5, implying a non-native disulphide bond between cysteines 1 and 6); fragments of mass 2353 and 2069 are not seen. Digestion products of the predominant disulphide-bonded species after folding of the E187K, ΔG197, D200G, D203N, S205P, D206E, E207Q, and E207K LR5 point mutants are within 3 a.m.u. of those predicted for the Phe 181–Glu 208 fragment, indicative of non-native disulphide bonds. In contrast, digestion products of the predominant disulphide-bonded species after folding of the D203G mutant of LR5 in buffer containing 1 mM $CaCl_2$ are within 3 a.m.u. of those predicted for the native disulphide bonded isomer.

NMR spectroscopy. NMR spectra were collected at protein concentrations of ~2 mM (pH 5.0) in 90% H_2O/10% D_2O in the absence of added buffer or salt. Data were acquired on a Bruker AMX 500-MHz spectrometer. ^1H spectra were collected at 30 °C using a spectral width of 6250 Hz and solvent presaturation, and were referenced to the carrier (4.76 p.p.m.). Data sets were defined by 2048 complex points and zero-filled once before Fourier transformation. Spectra were analysed using the Felix230 software package (Hare Research Inc.).

Calcium affinity of native LR5. Spectra were acquired on a Hitachi model F 4500 fluorescence spectrophotometer by excitation at 280 nm, monitoring emission at 358 nm. Incremental aliquots (10 μM) of an atomic absorption standard calcium solution (Sigma) were added to a solution of native LR5 (0.2 μM) in 10 mM PIPES buffer, pH 7.0, containing 100 mM NaCl and 0.1 mM EGTA. The concentration of unchelated calcium in the EGTA-containing solution was calculated using the program MaxChelator (MAXC v. 6.50), as described[38]. To obtain an estimate for K_d, the raw data were fit to equation 1, where F is the observed fluorescence, F_0 is the fluorescence of unliganded LR5, F_∞ is the fluorescence of fully liganded LR5, and $[Ca^{++}_{free}]$ is the unchelated calcium concentration, corrected for the ligand-bound fraction ($[Ca^{++}_{free}]$ = unchelated $[Ca^{++}]-[(F-F_0)/(F_\infty-F_0)]*[LR5]_{total}$).

Eq.1:

$$[(F-F_0)/(F_\infty-F_0)]/[Ca^{++}_{free}]=1/K_d-[(F-F_0)/(F_\infty-F_0)]/K_d$$

Stephen C. Blacklow[1,2] and Peter S. Kim[1]

[1]Howard Hughes Medical Institute, Whitehead Institute for Biomedical Research, and Department of Biology, Massachusetts Institute of Technology, Nine Cambridge Center, Cambridge, Massachusetts 02142, USA

[2]Department of Pathology, Brigham and Women's Hospital, 75 Francis Street, Boston, Massachusetts 02115, USA

Correspondence should be addressed to P.S.K.

Received 17 April; accepted 24 June1996.

correspondence

Acknowledgements We thank M. Milhollen for assistance in construction of the plasmid used for expression of LR5, B. Schulman for developing the Endo-Glu method for assignment of disulphide bond connectivities, and L. Wu for critical review of the manuscript. SCB is a recipient of a Howard Hughes Postdoctoral Fellowship for Physicians. This research was supported by an NIH grant from the Program of Excellence in Molecular Biology to R.D. Rosenberg.

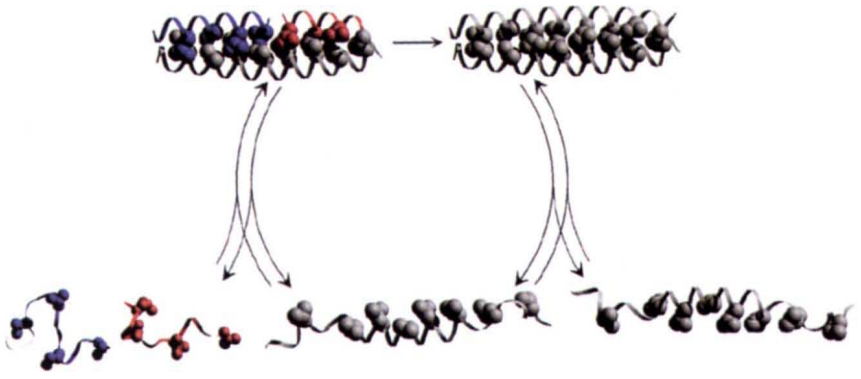
1. Brown, M.S. & Goldstein, J.L. *Science* **232**, 34–47 (1986).
2. Sudhof, T.C., Goldstein, J.L., Brown, M.S. & Russell, D.W. *Science* **228**, 815–22 (1985).
3. Yamamoto, T. *et al. Cell* **39**, 27–38 (1984).
4. Raychowdhury, R., Niles, J.L., McCluskey, R.T. & Smith, J.A. *Science* **244**, 1163–1165 (1989).
5. Herz, J. *et al. EMBO J.* **7**, 4119–4127 (1988).
6. Murdoch, A.D., Dodge, G.R., Cohen, I., Tuan, R.S. & Iozzo, R.V. *J. Biol. Chem.* **267**, 8544–8557 (1992).
7. Marazziti, D., Eggersten, G. Fey, G.H. & Stanley, K.K. *Biochemistry* **27**, 6529–6534 (1988).
8. Hobart, M.J., Fernie, B. & DiScipio, R.G. *Biochemistry* **32**, 6198–6205 (1993).
9. Bates, P., Young, J.A.T. & Varmus, H.E. *Cell* **74**, 1043–1051 (1993).
10. Hobbs, H.H., Brown, M.S. & Goldstein, J.L. *Human Mutation* **1**, 445–466 (1992).
11. Mahley, R.W. *Science* **240**, 622–630 (1988).
12. Bieri, S., Djordevic, J.T., Daly, N.L., Smith, R. & Kroon, P.A. *Biochemistry* **34**, 13059–13065 (1995).
13. Bieri, S., Djordevic, J.T., Jamshidi, N., Smith, R. & Kroon, P. *FEBS Letters* **371**, 341–344 (1995).
14. Hobbs, H.H., Russell, D.W., Brown, M.S. & Goldstein, J.L. *Annu Rev. Genet.* **24**, 133–170 (1990).
15. Leitersdorf, E., Van Der Westhuyzen, D., Coetzee, G.A. & Hobbs, H.H. *J. Clin. Invest.* **84**, 954–961 (1989).
16. Leitersdorf, E., Tobin, E.J., Davignon, J. & Hobbs, H.H. *J. Clin. Invest.* **85**, 1014–1023 (1990).
17. Schneider, W.J., Brown, M.S. & Goldstein, J.L. *Mol Biol Med* **1**, 353–67 (1983).
18. Yamamoto, T. Bishop, R.W. Brown, M.S. Goldstein, J.L. & Russell, D.W. *Science* **232**, 1230–1237 (1986).
19. Thomas, P.J., Qu, B.-H. & Pedersen, P.L. *TIBS* **20**, 456–459 (1995).
20. Thomas, P.J., Shenbagamurthi, P., Sondek, J., Hullihen, J.M. & Pedersen, P.L. *J. Biol. Chem.* **267**, 5727–5730 (1992).
21. Lomas, D.A., Evans, D.L.I., Finch, J.T. & Carrell, R.W. *Nature* **357**, 605–607 (1992).
22. Yu, M.-H., Lee, K.N. & Kim, J. *Nature Struct. Biology* **2**, 363–367 (1995).
23. Liu, X., Garriga, P. & Khorana, H.G. *Proc. Natl. Acad. Sci. USA* **93**, 4554–4559 (1996).
24. Wu, Y.-S., Bevilacqua, V.L.H. & Berg, J.M. *Chemistry and Biology* **2**, 91–97 (1995).
25. Babu, Y.S., Bugg, C.E. & Cook, W.J. *J. Mol. Biol.* **204**, 191–204 (1988).
26. Kawasaki, H. & Kretsinger, R.H. *Protein Profile* **1**, 343–350 (1994).
27. Handford, P.A. *et al. Nature* **351**, 164–167 (1991).
28. Rao, Z., Handford, P., Mayhew, M., Knott, V., Brownlee, G.G. & Stuart, D. *Cell* **82**, 131–141 (1995).
29. Wilson, C., Wardell, M.R., Weisgraber, K.H., Mahley, R.W. & Agard, D.A. *Science* **252**, 1817–1822 (1991).
30. Daly, N.L., Scanlon, M.J., Djordjevic, J.T., Kroon, P. & Smith, R. *Proc. Natl. Acad. Sci. USA* **92**, 6334–6338 (1995).
31. Daly, N.L., Djordevic, J.T., Kroon, P.A. & Smith, R. *Biochemistry* **34**, 14474–14481 (1995).
32. Staley, J.P. & Kim, P.S. *Protein Sci.* **3**, 1822–1832 (1994).
33. Kunkel, T.A., Roberts, J.D. & Zakour, R.A. *Methods Enzymol.* **154**, 367–382 (1987).
34. Sambrook, J., Fritsch, E.F. & Maniatis, T. *Molecular cloning: a laboratory manual*; Cold Spring Harbor Laboratory Press: Cold Spring Harbor, New York, 1989.
35. Studier, F.W., Rosenberg, A.H., Dunn, J.J., Dubendorff, J.W. *Methods Enzymol* **185**, 60–89 (1990).
36. Peng, Z.Y. & Kim, P.S. *Biochemistry* **33**, 2136–2141 (1994).
37. Blacklow, S.C., Lu, M. & Kim, P.S. *Biochemistry* **34**, 14955–14962 (1995).
38. Bers, D.M., Patton, C.W. & Nuccitelli, R. In *Methods in Cell Biology, Volume 40*, (Ed. Nuccitelli, R.) 3–29 (Academic Press: San Diego, 1994).
39. Wüthrich, K. *NMR of proteins and nucleic acids*. (Wiley-Interscience: New York, 1986).

picture story

Self-made molecules

Both serious scientific consideration and all-night collegiate hand-waving assert that self-replication is one of the hallmarks of life. It seems only reasonable to assume that at some point along the way to becoming living, inanimate matter passed through a stage in which a simple chemical system could replicate itself. At present, we are most comfortable with the idea that nucleic acids could have fulfilled this self-replicative role, due to their evident complementarity during normal replication, and, in some cases, their ability to act as enzymes. Now evidence has been found that peptides can catalyse their own formation as well (D.H. Lee, J.R.Granja, J.A. Martinez, K. Severin and E.R. Ghadiri *Nature* **382**, 425–528), broadening our notion of how autocatalytic self-replication could occur.

The elements of this autocatalytic system are the template—a 32 amino-acid α-helical peptide (grey) derived from GCN4—which catalyses its own synthesis from two peptide fragments of 17 and 15 amino acids (shown in blue and red in the figure). The template is designed to associate with itself (and also the fragments) through interhelical hydrophobic packing and electrostatic interactions. Tyrosines were engineered into the solvent-exposed surface to ease spectroscopic monitoring of the reaction. The ligation site of the two fragments also lies on the solvent-exposed surface. The 'food' for this self-replicating system requires careful preparation: the two fragments must be primed for the reaction. The C terminus of the electrophilic 17-mer (blue) is preactivated as a thiobenzyl ester, which is attacked by the N-terminal sulphydyl of the nucleophilic fragment (red) to form a thioester intermediate and then an amide bond. The reaction is specific (only 15% side products), and requires the ternary complex of template with both fragments (the two binary complexes were excluded as intermediates by testing templates that do not bind to one or the other fragment). The initial rate of formation of product increases with the square-root of the initial template concentration; further, product formation describes a sigmoidal curve, confirming that the process is autocatalytic.



Proteins are more versatile molecules than nucleic acids, and we can now add self-replication to their repertoire. (To those who have thought about prions, it may not be so surprising that a protein alone could catalyse its own formation, albeit in a different sense than the present case). Is this system just a fascinating molecular toy? Is it a way to study how living systems can originate? At the least, these results tell us that the necessary ingredients for self-replication, and perhaps the first makings of life, could be easier to attain than we had previously thought.

AF