



Published in final edited form as:

J Am Chem Soc. 2011 June 15; 133(23): 8810–8813. doi:10.1021/ja201581x.

Molecular Recognition of Insulin by a Synthetic Receptor

Jordan M. Chinai[†], Alexander B. Taylor^{‡,⊥}, Lisa M. Ryno[†], Nicholas D. Hargreaves[†], Christopher A. Morris[†], P. John Hart^{‡,⊥,Ⓢ}, and Adam R. Urbach^{*,†}

[†]Department of Chemistry, Trinity University, One Trinity Place, San Antonio, TX, 78212

[‡]Department of Biochemistry, South Texas Health Care System, The University of Texas Health Science Center at San Antonio, San Antonio, TX, 78229

[⊥]The X-ray Crystallography Core Laboratory, South Texas Health Care System, The University of Texas Health Science Center at San Antonio, San Antonio, TX, 78229

[Ⓢ]The Geriatric Research, Education, and Clinical Center, Department of Veteran's Affairs, South Texas Health Care System, The University of Texas Health Science Center at San Antonio, San Antonio, TX, 78229

Abstract

The discovery of molecules that bind tightly and selectively to desired proteins continues to drive innovation at the interface of chemistry and biology. This paper describes the binding of human insulin by the synthetic receptor cucurbit[7]uril (Q7) *in vitro*. Isothermal titration calorimetry and fluorescence spectroscopy experiments show that Q7 binds to insulin with an equilibrium association constant of $1.5 \times 10^6 \text{ M}^{-1}$ and with 50-100-fold selectivity versus proteins that are much larger but lack an N-terminal aromatic residue, and >1000-fold selectivity versus an insulin variant lacking the N-terminal phenylalanine (Phe) residue. The crystal structure of the Q7•insulin complex shows that binding occurs at the N-terminal Phe residue and that the N-terminus unfolds to enable binding. These findings suggest that site-selective recognition is based on the properties inherent to a protein terminus, including the unique chemical epitope presented by the terminal residue and the greater freedom of the terminus to unfold, like the end of a ball of string, to accommodate binding. Insulin recognition was predicted accurately from studies on short peptides and exemplifies an approach to protein recognition by targeting the terminus.

Living systems present a highly complex array of molecular surfaces that vary in size, shape, polarity, organization and solvation, and yet each molecule finds its target destination with high fidelity. The weak intermolecular forces (e.g., coulombic, dispersive, and solvophobic) that direct these events are well understood in principle.¹⁻³ In practice, however, the design of synthetic compounds to recognize desired protein targets presents persistent challenges en route to pharmaceuticals, medical diagnostics, and other tools for biochemistry and chemical biology. So-called “rational” approaches to this problem use detailed knowledge of the structure of a target to aid in the design of a complementary molecule.⁴⁻⁶ Although high-resolution structures are available for a number of protein targets, the precise structure of the surface of a protein and the surrounding solvent remains difficult to predict from the genetically encoded sequence of amino acids or from homology modeling. Therefore, general design principles for sequence-based recognition of folded proteins do not currently exist. By contrast, numerous groups have reported on the sequence-

* to whom correspondence should be addressed: aurbach@trinity.edu.

Supporting Information Available: Experimental details, thermodynamic binding data, X-ray crystallographic details, circular dichroism spectra.

specific recognition of short peptides in aqueous solution by synthetic receptors.⁷⁻¹⁷ Short peptides are more practical than proteins for structure-activity studies of molecular recognition because they are less expensive, are easier to modify, and in many cases their structures are predictably folded/unfolded. Therefore, strategies for translating known approaches of peptide recognition into protein recognition would be exceptionally valuable. Here we show that the synthetic receptor cucurbit[7]uril (Q7, Figure 1) binds to the N-terminal phenylalanine (Phe) of human insulin, and that this recognition enables Q7 to bind more tightly to insulin than to proteins lacking an N-terminal aromatic residue. Equilibrium binding studies and X-ray crystallography reveal the molecular basis for recognition and show how the properties of a protein terminus are well suited to site-selective recognition.

Q7 is the seven-membered macrocycle in the cucurbit[n]uril (Qn) family of water-soluble, synthetic receptors, which have gained importance for their capacity to bind small guests very tightly and selectively in aqueous solution.¹⁸⁻²⁰ The nonpolar interior and carbonyl-lined portals of Qn receptors work in concert to bind organic cations by including the hydrophobic portion of the guest within the apolar cavity and stabilizing the cationic group(s) with the portal oxygens. This cooperation between cavity and portal has proved ideal for binding to short peptides with an N-terminal phenylalanine (Phe), tryptophan (Trp), or tyrosine (Tyr), in a sequence-specific fashion.^{14-17,21} In focused structure-activity studies with Q7 and Q8,^{15,17} all peptides had a single aromatic group and an N-terminal ammonium group, but only those with an N-terminal aromatic residue, which fixes both groups in close proximity, were bound selectively by Q7 and Q8. Therefore, the terminus provided a binding site that is structurally unique at the level of a single-residue and thus known from the sequence of amino acids. On the basis of these studies we predicted that a Qn receptor should bind selectively to an N-terminal aromatic residue *versus all other surface-exposed aromatic residues on a folded protein* and thus provide a mechanism for site-specific protein recognition. In the present study, human insulin was identified as a desirable model on which to test this hypothesis because it has a rare Phe residue at the N-terminus of its B-chain (i.e., Phe^{B1}), it has been thoroughly characterized, and it is stably folded under ambient conditions.²²

Isothermal titration calorimetry (ITC) experiments revealed that Q7 binds to recombinant human insulin **1** in a 1:1 stoichiometric ratio with an equilibrium association constant (K_a) value of $1.5 \times 10^6 \text{ M}^{-1}$ (Figure 2). Although we anticipated binding at the Phe^{B1} position, insulin also has on its surface two Phe and four Tyr residues, all of which are possible binding sites. The series of compounds **2-5** (Figure 1) was designed to probe the location of binding. Protein **2** is a variant of human insulin with glutamic acid (Glu) at positions B1 and B27 (i.e., Glu^{B1}, Glu^{B27}).^{23,24} Peptides **3-5** are models of the N-terminal Phe (**3**), non-terminal Phe (**4**), and non-terminal Tyr (**5**) residues on the surface of insulin.

ITC measurements (Figure 2 and Supporting Information) showed no measurable binding of Q7 to variant insulin **2**, which has no N-terminal Phe, and thus a decrease in association constant by more than three orders of magnitude compared to native insulin **1**. This result makes a strong case for Phe^{B1} as the site of binding. Positive control peptide **3** ($K_a = 2.8 \times 10^6 \text{ M}^{-1}$), which has an N-terminal Phe, binds to Q7 with similar affinity as native insulin **1**. Peptides **4** ($2.2 \times 10^4 \text{ M}^{-1}$) and **5** ($2.7 \times 10^3 \text{ M}^{-1}$) were bound less tightly than peptide **3** by two and three orders of magnitude, respectively, showing that non-terminal Phe and Tyr residues are possible, albeit weak, sites of binding for Q7. Therefore, we were surprised that Q7 showed no measurable affinity for variant insulin **2**, which suggests that the two non-terminal Phe and four non-terminal Tyr residues on the surface of **2** are significantly less accessible to binding than the analogous residues in peptides **4** and **5**.

We sought structural evidence to support the ITC data and to explore a molecular basis for recognition. Toward this end, we determined the crystal structure of the Q7•**1** complex (Figure 3). The asymmetric unit of the crystal (Figure 3a) contains two insulin molecules (**1a** and **1b**) and one molecule of Q7, which is associated with insulin **1a** at the N-terminus of the B chain. Akin to previously reported structures of Q8 in complex with Phe-Gly-Gly and Trp-Gly-Gly,¹⁵ the Q7•insulin structure shows the aromatic sidechain of Phe^{B1} included within the cavity of Q7. In addition, Val^{B2} and Asn^{B3} contribute weak stabilizing interactions by contacting atoms of the Q7 exterior. Overall, the binding of Q7 buries approx. 200 Å² of insulin's solvent-accessible surface area.²⁵ The presence of Q7-bound insulin **1a** and unbound insulin **1b** in the asymmetric unit of the crystal allows us to study perturbations to the structure of insulin upon binding to Q7. Overall, the structures of **1a** and **1b** are strikingly similar, with a root mean square deviation (RMSD) of 0.37 Å over 35 alpha carbons. The largest deviations are observed at the B1-B4 residues of **1a**, which unfold from the surface of the protein to accommodate Q7 (Figure 3b).

These results corroborate the calorimetric data and provide a structural basis for recognition at the N-terminus of the B chain of insulin. As observed with small peptides,¹⁵ stable binding requires the simultaneous inclusion of the hydrophobic sidechain of Phe^{B1} within the cavity of Q7 and interaction of the N-terminal ammonium group with the portal oxygens of Q7. The Phe^{B1} residue therefore presents a unique binding epitope comprising the aromatic sidechain and the N-terminal ammonium group, which cannot exist anywhere else on the protein. Moreover, the ability of the insulin N-terminus to partially unfold may be critical to binding in the sterically demanding context of a folded protein. The singular connection of the terminal residue to the protein, like the end of a ball of string, facilitates the adoption of a fully solvent-exposed, random-coil structure, akin to short peptides. This result is likely not unique to insulin because protein termini are typically solvent-exposed, more so than charge alone would predict.²⁶

Based on the mechanism of binding established here, we expected that Q7 would not exhibit high affinity binding to proteins lacking an N-terminal aromatic residue. A fluorescence indicator displacement assay²⁷ was used to test this hypothesis. We chose acridine orange because it has relatively little nonspecific association with proteins, it is effective at neutral pH, and its K_a value for Q7 ($2.0 \times 10^5 \text{ M}^{-1}$) is similar yet lower than that of insulin.²⁸ The emission of acridine orange increases when bound to Q7. Therefore, upon introducing a competitive analyte to a solution containing the Q7•acridine orange complex, the displacement of acridine orange from Q7 yields a decrease in fluorescence intensity that indicates the binding of analyte (Figure 4a).

The fluorescence response for insulin was compared to compounds **2-5**, the peptide Gly-Gly-Gly, and a series of commercially available blood proteins, including bovine serum albumin (BSA), human immunoglobulin G (IgG), and bovine carbonic anhydrase (BCA) (Figure 4b). We observed substantial quenching (55-80%) with insulin **1** and peptide **3** (positive control), but insignificant quenching for the other analytes (1-5%), which lack an aromatic terminus. This result demonstrates that a single residue can enable the selective binding of insulin versus other proteins by at least 50-100-fold in K_a²⁹ under the conditions used in this study. We note explicitly, however, that insulin recognition *in vivo* would be precluded by (i) its relatively high dissociation constant (0.67 μM) compared to the sub-nM concentrations of insulin in blood and saliva, (ii) the presence of weak binding proteins (e.g., albumin) at sufficiently high concentrations to effectively compete for binding,³⁰⁻³² and (iii) the possibility of competition by other peptides and proteins with N-terminal aromatic groups. This paper does not claim to address these challenges pertaining to *in vivo* applications.

This paper describes an approach to protein recognition that exploits the unique properties of a protein terminus to achieve single-site binding that is predictable from the sequence of amino acids. Unlike small molecules that bind inside protein cavities,⁵ and unlike natural and synthetic receptors that bind to large areas on the surfaces of proteins,⁴ Q7 binds to insulin by incorporating the terminal residue within its cavity. When the terminus unravels from the surface it becomes fully solvent exposed, thus allowing Q7 to bind to insulin in the same manner as to a short peptide. This motif is extraordinarily minimal, comprising a small and structurally simple receptor, a single amino acid residue, and a binding interface of $\sim 200 \text{ \AA}^2$. This approach to protein recognition differs from existing methods, such as genetically engineering polyhistidine (His-tag) or glutathione *S*-transferase (GST) fusion proteins into target proteins for affinity chromatography,³³ or chemically modifying proteins with biotin or aminomethylferrocene groups for capture using avidin or Q7,³⁴⁻³⁶ respectively. In the current study, Q7 binds directly to the native protein sequence, not an engineered or chemically modified sequence,³⁷ although we consider these techniques to be complementary. We believe this approach will facilitate the design of receptors for other proteins as well as the development of techniques for single-site labeling of proteins and for sensing and separating proteins according to the identity of the terminal residue.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

This paper is dedicated to Prof. Jonathan L. Sessler on the occasion of his 55th birthday. Support for this work from the National Science Foundation (CHE-0748483, ARU) and Welch Foundation (W-1640, ARU and AQ-1399, PJH) is gratefully acknowledged. We thank Novo Nordisk for generous donation of the Glu^{B1}, Glu^{B27} variant insulin **2**. The circular dichroism spectropolarimeter was provided by a grant from the National Science Foundation (DBI-0718766, ARU). The fluorescence spectrometer and titration calorimeter were provided by a grant to Trinity University by the W. M. Keck Foundation. LMR had a Jean Dreyfus Boissevain Undergraduate Scholarship for Excellence in Chemistry from the Camille and Henry Dreyfus Foundation. This work is based on research conducted at the Advanced Photon Source on the Northeastern Collaborative Access Team (NE-CAT) beamlines, which are supported by award RR-15301 from the National Center for Research Resources at the National Institutes of Health. Use of the Advanced Photon Source is supported by the U. S. Department of Energy, Office of Basic Energy Sciences, under Contract No. DE-AC02-06CH11357. We thank Dr. Jonathan P. Schuermann of NE-CAT for collecting the data for the Q7•**1** complex. Support for the *X-ray Crystallography Core Laboratory* at the University of Texas Health Science Center by the Institutional Executive Research Council and the San Antonio Cancer Institute is gratefully acknowledged.

References

1. Schneider HJ. *Angew Chem Intl Ed.* 2009; 48:3924–3977.
2. Williams DH, Stephens E, O'Brien DP, Zhou M. *Angew Chem Intl Ed.* 2004; 48:6596–6616.
3. Meyer EA, Castellano RK, Diederich F. *Angew Chem Intl Ed.* 2003; 42:1210–1250.
4. Giralt, E.; Peczu, M.; Salvatella, X., et al., editors. *Protein surface recognition approaches for drug design*. John Wiley & Sons; West Sussex: 2001.
5. Böhm, HJ.; Schneider, G., et al., editors. *Protein-ligand interactions: from molecular recognition to drug design*. Wiley-VCH; Weinheim:
6. Krishnamurthy VM, et al. *Chem Rev.* 2008; 108:946–1051. [PubMed: 18335973]
7. Peczu MW, Hamilton AD. *Chem Rev.* 2000; 100:2479–2494. [PubMed: 11749292]
8. Breslow R, Yang Z, Ching R, Trojandt G, Odobel F. *J Am Chem Soc.* 1998; 120:3536–3537.
9. Hossain MA, Schneider HJ. *J Am Chem Soc.* 1998; 120:11208–11209.
10. Wagner H, Still WC, Chen CT. *Science.* 1998; 279:851–853. [PubMed: 9452382]
11. Tashiro S, Tominaga M, Kawano M, Therrien B, Ozeki T, Fujita M. *J Am Chem Soc.* 2005; 127:4546–4547. [PubMed: 15796499]

12. Schmuck C. *Coord Chem Rev.* 2006; 250:3053–3067.
13. Wehner M, Janssen D, Schafer G, Schrader T. *Eur J Org Chem.* 2006:138–153.
14. Bush ME, Bouley ND, Urbach AR. *J Am Chem Soc.* 2005
15. Heitmann LM, Taylor AB, Hart PJ, Urbach AR. *J Am Chem Soc.* 2006; 128:12574–12581. [PubMed: 16984208]
16. Rekharsky MV, Yamamura H, Inoue C, Kawai M, Osaka I, Arakawa R, Shiba K, Sato A, Ko YH, Selvapalam N, Kim K, Inoue Y. *J Am Chem Soc.* 2006; 128:14871–14880. [PubMed: 17105297]
17. Rekharsky MV, Yamamura H, Ko YH, Selvapalam N, Kim K, Inoue Y. *Chem Comm.* 2008:2236–2238. [PubMed: 18463751]
18. Lee JW, Samal S, Selvapalam N, Kim HJ, Kim K. *Acc Chem Res.* 2003; 36:621–630. [PubMed: 12924959]
19. Lagona L, Mukhopadhyay P, Chakrabarti S, Isaacs L. *Angew Chem Int Ed.* 2005; 44:4844–4870.
20. Isaacs L. *Chem Comm.* 2009:619–629. [PubMed: 19322405]
21. Tian F, Cziferszky M, Jiao D, Wahlström K, Geng J, Scherman OA. *Langmuir.* 2011; 27:1387–1390. [PubMed: 21194231]
22. Federwisch, M.; Dieken, ML.; De Meyts, P., et al., editors. *Insulin and Related Proteins – Structure to Function and Pharmacology.* Kluwer Academic Publishers; Dordrecht, The Netherlands: 2002.
23. Uversky VN, Garriques LN, Millett IS, Frokjaer S, Brange J, Doniach S, Fink AL. *J Pharmaceut Sci.* 2003; 92:847–858.
24. We expect the additional mutation of Thr^{B27} to Glu to have no impact on Q7 binding because prior work has shown that Qn's do not bind to Thr residues.^{38,39} Moreover, the circular dichroism (CD) spectrum of 2 is essentially identical to that of 1 (see Supporting Information), which shows that the mutations do not influence the folded structure of insulin.
25. Brunger AT. *Acta Cryst D.* 1998; 54:905–921. [PubMed: 9757107]
26. Jacob E, Unger R. *Bioinformatics.* 2006; 23:e225–e230. [PubMed: 17237096]
27. Nguyen BT, Anslyn EV. *Coord Chem Rev.* 2006; 250:3118–3127.
28. Shaikh M, Mohanty J, Singh PK, Nau WM, Pal H. *Photochem Photobiol Sci.* 2008; 7:408–414. [PubMed: 18385882]
29. Calculated from the fraction of fluorescence quenching observed under these competitive equilibrium conditions.
30. Bhasikuttan AC, Mohanty J, Nau WM, Pal H. *Angew Chem Int Ed.* 2007; 46:4120–4122.
31. Shaikh M, Mohanty J, Bhasikuttan AC, Uzunova VD, Nau WM, Pal H. *Chem Comm.* 2008:3681–3683. [PubMed: 18665298]
32. Lei W, Jiang G, Zhou Q, Zhang B, Wang X. *Phys Chem Chem Phys.* 2010; 12:13255–13260. [PubMed: 20824256]
33. Hochuli E, Bannwarth W, Döbeli H, Gentz R, Stüber D. *Nature Biotech.* 1988; 6:1321–1325.
34. Hwang I, Baek K, Jung M, Kim Y, Park KM, Lee DW, Selvapalam N, Kim K. *J Am Chem Soc.* 2007; 129:4170–4171. [PubMed: 17373802]
35. Young JF, Nguyen HD, Yang L, Huskens J, Jonkheijm P, Brunsveld L. *Chembiochem.* 2010; 11:180–183. [PubMed: 19937592]
36. Lee DW, Park KM, Banerjee M, Ha SH, Lee T, Suh K, Paul S, Jung H, Kim J, Selvapalam N, Ryu SH, Kim K. *Nature Chem.* 2011; 3:154–159. [PubMed: 21258389]
37. Nguyen HD, Dang DT, van Dongen JLJ, Brunsveld L. *Angew Chem Int Ed.* 2010; 49:895–898.
38. Rajgariah P, Urbach AR. *J Incl Phenom Macrocycl Chem.* 2008; 62:251–254.
39. Cong H, Tao LL, Yu YH, Yang F, Du Y, Xue SF, Tao Z. *Acta Chim Sin.* 2006; 64:989–996.

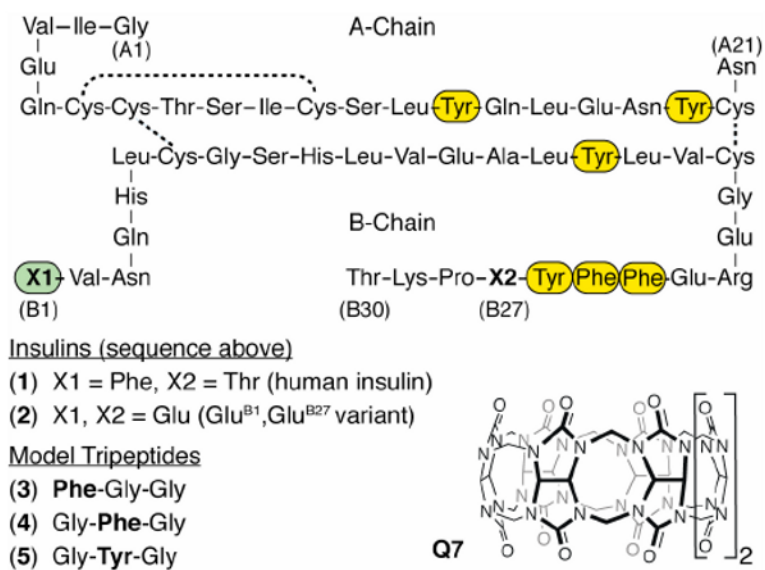


Figure 1.
 Amino acid sequences of the insulins and model peptides used in this study, and the chemical formula of Q7.

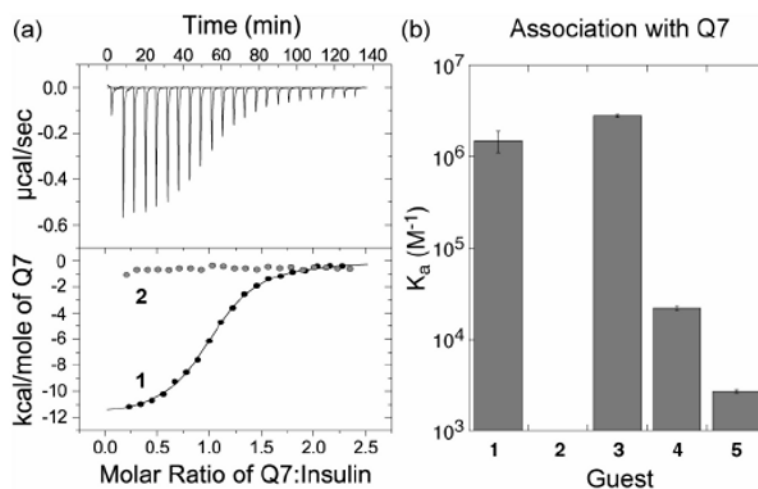


Figure 2.

(a) Representative ITC data for the titration of Q7 to native human insulin **1**. The top plot of power versus time was integrated to yield the bottom plot of molar enthalpy versus the ratio of Q7:insulin. The gray data points in the bottom plot are for the analogous experiment on variant insulin **2**, showing no measurable binding. (b) Bar plot of the equilibrium association constants of analytes **1** – **5** in complex with Q7. Error bars are standard deviations from at least three measurements. The ITC technique has a practical lower limit of 10^3 M^{-1} .

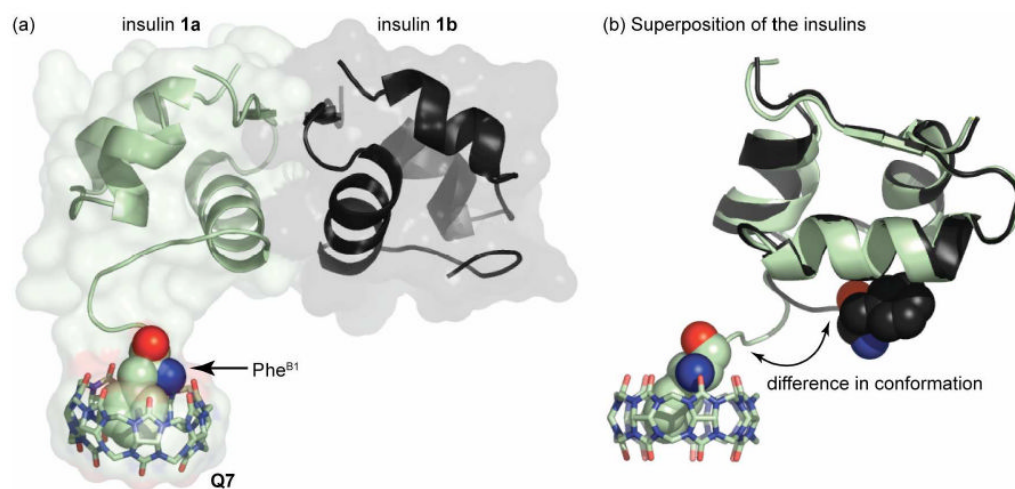


Figure 3. Crystal structure of the Q7• **1** complex. (a) Asymmetric unit of the crystal, with the two molecules of insulin, **1a** (green) and **1b** (black), rendered as ribbons and the associated Q7 as sticks. The solvent-accessible surface is shown as semi-transparent to reveal the overall shape of the complex. The Phe^{B1} position of **1a**, which is bound inside Q7, is rendered as space-filling, with carbon in green, oxygen in red, and nitrogen in blue; the N-terminal nitrogen is the blue sphere. (b) Superposition of the two insulins illustrates the strong homology between **1a** and **1b** with clear divergence at the N-terminus of the B-chain. In this image, the Phe^{B1} position of both insulins is rendered as space-filling for comparison.

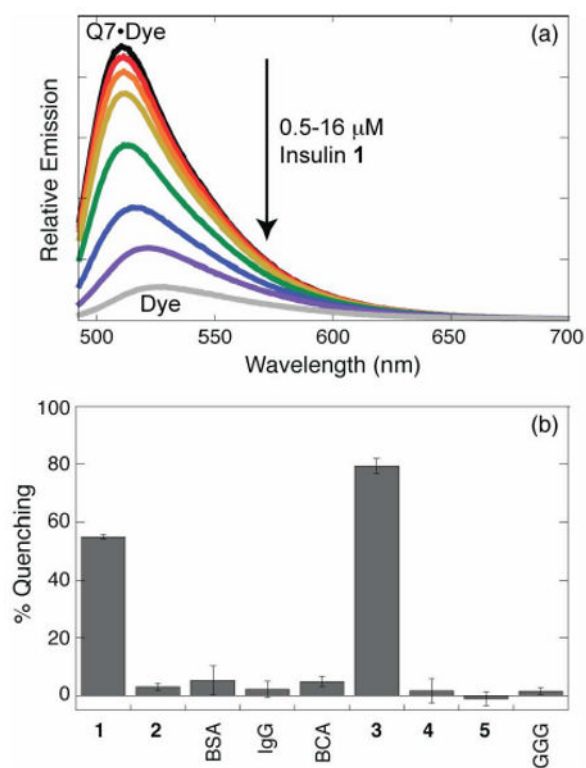


Figure 4. Fluorescent indicator displacement assay for insulin. (a) Decrease in emission intensity upon titrating insulin at 0, 0.5, 1.0, 2.0, 4.0, 8.0, and 16.0 μM into a mixture of Q7 (6 μM) and acridine orange (8 μM). (b) Fluorescence quenching for a series of proteins and peptides at 4 μM in a mixture of 4 μM Q7 and 6 μM acridine orange, showing selective binding of native insulin **1** and positive control peptide **3**. Error bars are standard deviations from at least three measurements. BSA, bovine serum albumin; IgG, immunoglobulin G; BCA, bovine carbonic anhydrase; GGG, peptide Gly-Gly-Gly.