Design of efficient molecular organic light-emitting diodes by a high-throughput virtual screening and experimental approach

Rafael Gómez-Bombarelli[1], Jorge Aguilera-Iparraguirre[1], Timothy D. Hirzel[1], David Duvenaud[2], Dougal Maclaurin[2], Martin A. Blood-Forsythe[1], Hyun Sik Chae[3], Markus Einzinger[4], Dong-Gwang Ha[5], Tony Wu[4], Georgios Markopoulos[6], Soonok Jeon[7], Hosuk Kang[7], Hiroshi Miyazaki[7], Masaki Numata[7], Sunghan Kim[7], Wenliang Huang[6], Seong Ik Hong[3], Marc Baldo[4], Ryan P. Adams[2] and Alán Aspuru-Guzik[1]*

Virtual screening is becoming a ground-breaking tool for molecular discovery due to the exponential growth of available computer time and constant improvement of simulation and machine learning techniques. We report an integrated organic functional material design process that incorporates theoretical insight, quantum chemistry, cheminformatics, machine learning, industrial expertise, organic synthesis, molecular characterization, device fabrication and optoelectronic testing. After exploring a search space of 1.6 million molecules and screening over 400,000 of them using time-dependent density functional theory, we identified thousands of promising novel organic light-emitting diode molecules across the visible spectrum. Our team collaboratively selected the best candidates from this set. The experimentally determined external quantum efficiencies for these synthesized candidates were as large as 22%.

The availability of unprecedented computational resources and novel algorithms is opening pathways to solve problems that were previously intractable due to their large scale. For this reason, computational tools can become a disruptive instrument for molecular discovery. Chemical space is infinite for practical purposes[1], but many important and desirable properties of materials are predictable from simulation. However, computers alone cannot meet this challenge, no matter how powerful. It is only with experimental guidance and collaboration that a computer-driven approach becomes a natural fit for this task. High-throughput virtual screening (HTVS) combines quantum chemical calculations and cheminformatics methods to search molecular space for leads[2–4]. Advised by computational predictions, experimental efforts are focused on only the most promising molecular candidates. Recent applications of this approach include screening for both inorganic[5,6] and organic materials in application areas such as electrolytes for flow batteries[7,8] and photovoltaics[9–11]. However, it is not often that HTVS predictions are confirmed experimentally, due to factors such as access to experimental partners, synthetic difficulty, poor predictive power, and instability of candidate molecules.

The emissive layers of organic light-emitting diode (OLED) devices are made from electroluminescent molecules. Because of their high efficiency and superior colour properties, OLEDs have wide application in small displays and the OLED technology has the potential to become an attractive solution for larger displays such as televisions, flexible displays, or lighting applications. Traditional fluorescent emitters rely on the quantum-mechanically allowed singlet–singlet transition, and therefore can harvest only one quarter of the electron–hole recombination pairs formed in the device. Current phosphorescent OLED technology leverages spin–orbit coupling due to heavy atoms[12], most often iridium. This makes the triplet state emissive, and also harvests the singlet excitons through these emissive triplets[13]. However, iridium is one of the least abundant elements, and thus phosphorescent OLEDs face high materials cost. Additionally, OLEDs face challenges in efficiency and stability. In particular, deep-blue emitters magnify these problems due to the large amount of energy stored in the excited state, around 3 eV. The promises and challenges of OLED technology and materials have sparked significant research interest, as detailed in recent reviews[14,15].

One particularly promising approach to circumvent the high cost of phosphorescent OLED materials is thermally activated delayed fluorescence (TADF)[16–18]. In TADF, non-emissive triplet states are harvested via thermal fluctuations that repopulate the emissive singlet state through a process known as reverse intersystem

[1]Department of Chemistry and Chemical Biology, 12 Oxford Street, Harvard University, Cambridge, Massachusetts 02138, USA. [2]John A. Paulson School of Engineering and Applied Sciences, 33 Oxford Street, Harvard University, Cambridge, Massachusetts 02138, USA. [3]Samsung Research America, 255 Main Street, Suite 702, Cambridge, Massachusetts 02142, USA. [4]Department of Electrical Engineering and Computer Science, 77 Massachusetts Avenue, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139, USA. [5]Department of Materials Science and Engineering, 77 Massachusetts Avenue, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139, USA. [6]Department of Chemistry, 77 Massachusetts Avenue, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139, USA. [7]Samsung Advanced Institute of Technology, Samsung Electronics Co., Ltd., 130 Samsung-ro, Yeongtong-gu, Suwon-si, Gyeonggi-do 16678, Korea. *e-mail: aspuru@chemistry.harvard.edu
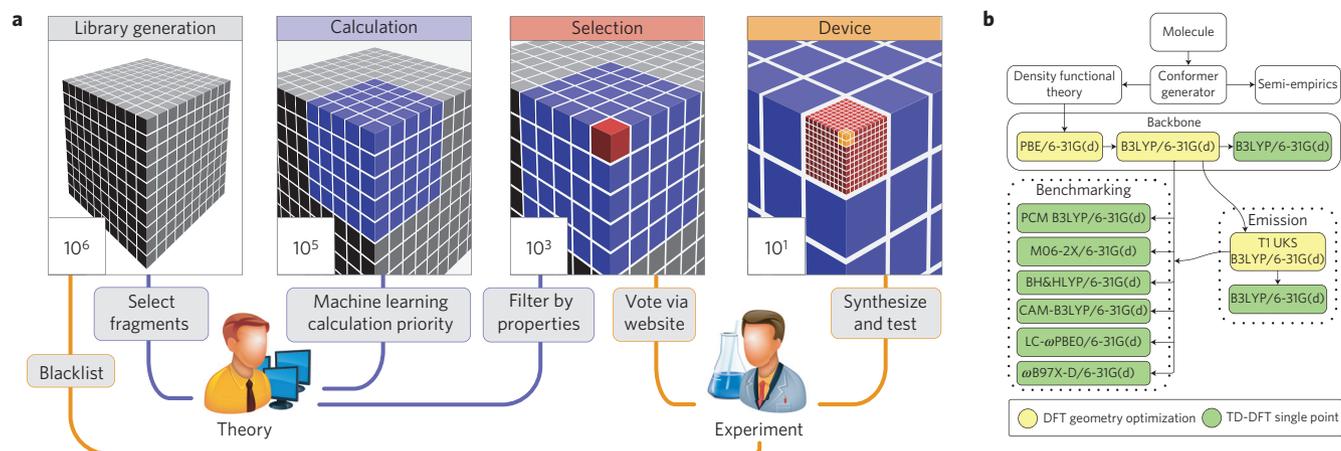
Figure 1 | Discovery pipeline. a, Diagram of the collaborative discovery approach: the search space decreases by over five orders of magnitude as the screening progresses. The cubes represent the size of the chemical space considered at any given stage of the process. The distinct screening stages, from left to right, involve different theoretical and computational approaches as well as experimental input and testing. **b**, Dependency tree for the quantum chemistry calculations employed in this study. The calculations labelled as backbone were performed for all analysed molecules, leading compounds were also characterized using the methods labelled emission, and the benchmarking calculations were used to assess predictive power.

crossing (RISC). This process occurs most efficiently with a small energy gap between the lowest singlet (S_1) and triplet (T_1) states. TADF emitters feature two characteristic excited-state decay times. The prompt component, on the order of nanoseconds, corresponds to the draining of S_1, and the slower delayed component, on the order of microseconds, corresponds to the RISC followed by fluorescent emission. The timescale of the prompt component is typical of conventional fluorescent emitters, while the delayed component lifetime can improve that of phosphorescent materials, which typically plateau around 1 µs (ref. 19). Triplet harvesting efficiencies close to 100% have been obtained through TADF[20,21] even in undoped devices[22]. No heavy atoms are required for TADF, and thus it presents a potentially efficient, cost-effective OLED technology.

There is a growing literature where small families of TADF compounds are explored experimentally[23–28], or purely computationally at a slightly larger scale[29,30]. Here, we report a large-scale computer-driven search for novel TADF OLED emitters, with a special focus on blue TADF emitters. This work employed a collaborative approach encompassing computational quantum chemistry, machine learning, organic synthesis, device fabrication and testing, and the expertise of industrial partners. This integrated cooperative effort aimed massive HTVS capabilities towards the successful discovery of novel materials. Figure 1a summarizes our screening approach, highlighting the roles of theory and experiment as the candidate count is reduced by five orders of magnitude.

In the following sections, we will describe the different screening stages that led to the discovery of OLED emitters with external quantum efficiency (EQE) values of up to 22%.

Library generation

The molecular search space was a virtual chemical library: a collection of graph representations of molecules stored in a database. Our library of more than 1.6 million candidates was created using in-house software that relies on the RDKit package[31]. Our software accepts a pool of starting fragments and provides a constrained combinatorial enumeration. The growth of the library took place in several generations, driven by chemical intuition, quantum simulation and experimental results.

OLED molecules with TADF character must contain a donor and an acceptor moiety, since only a charge-transfer excitation can exhibit a low enough singlet–triplet gap (ΔE_{ST}) for efficient thermal RISC. Thus, our candidate libraries obeyed a simple recipe:

donor-(bridge)$_x$–acceptor where x is between 0 and 2. This structure attempts to minimize the spatial overlap between the highest occupied molecular orbital (HOMO) and the lowest unoccupied molecular orbital (LUMO). Some of the key aspects of our library enumeration were control of symmetry in molecular substitution patterns and chemically informed substitution schemes and binding modes. Molecular size was limited by the requirement for vapour processing, with molecular mass cutoff at 1,100 g mol^{-1}.

A computational pre-screening of starting fragments was also performed for favourable optical properties. Since the lowest triplet excited state needs to be able to undergo RISC into a blue singlet, local triplet excitations were required to lie above a cutoff chosen at 2.60 eV. Moieties were flagged as donors, acceptors or both, using cutoffs of $E_{HOMO} > -6.5$ eV and $E_{LUMO} < -1.0$ eV. In aggregate, we combined 110 donor, 105 acceptor and 7 bridge moieties to generate over 1.6 million molecules (the fragments used to generate the library are available in section 3 of the Supplementary Information). To prevent chemically unstable patterns from emerging in the combinatorial growth process, a blacklist of disallowed substructures was compiled. Ertl and Schuffenhauer's synthetic accessibility (SA) score was used to estimate the synthetic accessibility of candidate molecules[32].

Quantum chemical calculations

The quantum chemical calculation scheme (Fig. 1b) aimed to strike a balance between predictive accuracy, robustness and computational cost. Although OLED emitters function within a device, their surrounding molecular environment is not as structured as other optoelectronic systems such as crystalline organic transistors or organic photovoltaics. Thus, environment effects can often be corrected via experimental calibration or approximated with minimal computational cost using implicit solvent methods.

Emission colour was estimated using time-dependent density functional theory (TD-DFT) vertical absorption energies, which are more affordable than emission wavelength calculations, since the later require optimized excited-state geometries. We obtained excellent accuracy by calibrating vertical absorption to photoluminescent emission colour in toluene solution (Fig. 2a). The effect of excited-state relaxation on both the T_1 and S_1 surfaces was also considered by conducting geometry optimization on the respective unrestricted density functional theory (UDFT) and TD-DFT potential energy surfaces.
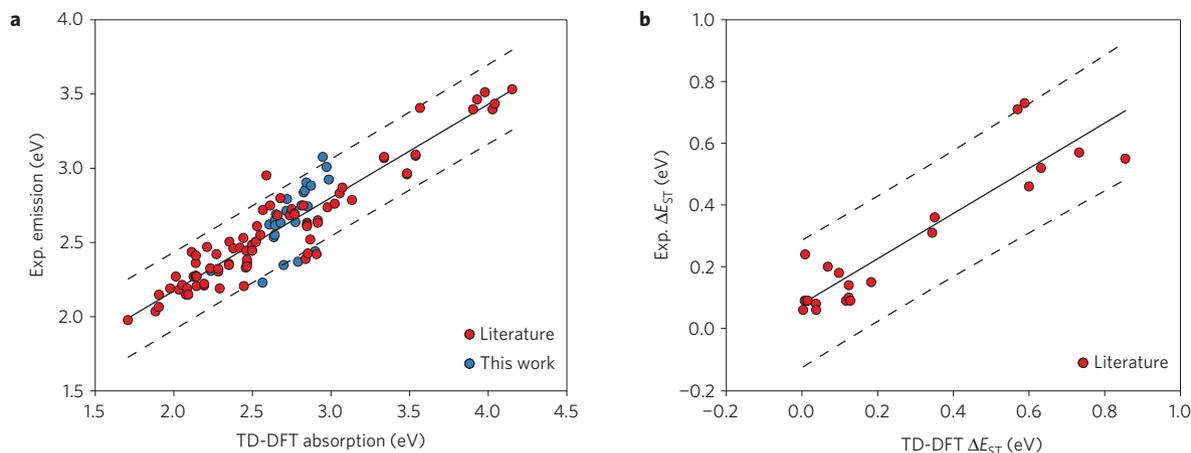
Figure 2 | Experiment–theory calibration. a, TD-DFT/B3LYP/6-31G(d) vertical absorption against photoluminescent emission maximum in toluene solution. Data in red represent measurements reported in the literature, while data in blue correspond to compounds synthesized in this screening work. The lines indicate a linear fit against the literature data (solid) with 95% confidence bounds (dashed). **b**, Experiment–theory comparison of TD-DFT/B3LYP/6-31G(d) singlet–triplet gap against experimental values, determined via the thermal activation energy method, measured in frozen toluene solution. The lines indicate a linear fit against the literature data (solid) with 95% confidence bounds (dashed).

To maximize the EQE in a TADF emitter it is necessary to boost the quantum yield of both the prompt and delayed components. For this, the rate of radiative emission of both the singlet and triplet excitons needs to be maximized, via fluorescence and delayed fluorescence, respectively, while minimizing the competing non-radiative pathways. Whereas the non-radiative decay rates are theoretically and computationally challenging to predict, and therefore unsuitable for HTVS, simple models can provide useful estimates for the radiative rates of both processes.

The radiative rate of fluorescence is proportional to the oscillator strength (f) of the $S_1 \rightarrow S_0$ transition, which is proportional to the square of the transition dipole moment $\int \psi_a(r) r \psi_b(r) \mathrm{d}^3 r$, where ψ_a and ψ_b are the wavefunctions of the initial and final state respectively. f is easily obtainable from excited-state quantum chemical calculations. The rate of TADF is more complex, since, from a molecular perspective, two essentially opposing quantities need to be optimized. On the one hand, f is maximized, to speed up the draining of S_1. On the other hand, ΔE_{ST} is minimized, since it is exponentially related to the rate of RISC. In the case of singlet and triplet excitations involving the same orbitals, ΔE_{ST} corresponds to the exchange integral $\int \psi_a(r_1) \psi_b(r_2)(1/r_{12}) \psi_a(r_2) \psi_b(r_1) \mathrm{d}^3 r_1 \mathrm{d}^3 r_2$, which directed the choice of molecular fragments with small frontier orbital overlap (ψ_a and ψ_b are the wavefunctions of electron 1 and 2, r_i is their position in space and r_{12} is the interelectronic distance).

For the purpose of evaluating candidates, it is useful to have a single-parameter figure of merit for TADF character. An upper bound on the delayed fluorescence rate constant (k_{TADF}) was estimated for this goal. Since the rate constants of RISC (k_{RISC}) and non-radiative decay (k_{nr}^T and k_{nr}^S for triplet and singlet, respectively) are inaccessible to HTVS calculations, an approximate value for the TADF rate constant was derived under the assumptions that $k_{nr}^T \approx k_{nr}^S \approx 0$, and the intersystem crossing quantum yield from S_1 to T_1 is unity[33]. The accuracy of these approximations in a variety of regimes is discussed in section 1.3 of the Supplementary Information.

The pre-equilibrium approximation was assumed in a regime where the thermal equilibrium between T_1 and S_1 is fast, and employed a Boltzmann factor weighted by the ratio of triplet to singlet states. The radiative fluorescent lifetime of $S_1(k_f)$ is estimated using a standard approach[34], and the product of k_f and the Boltzmann factor allows an estimated upper bound on the k_{TADF} rate constant (equation (1)), within the accuracy of TD-DFT. In the provided k_{TADF} equation e is the elementary charge, n is the refractive index of the medium, ϵ_0 is the vacuum permittivity, m_e is the electron

rest mass, c is the speed of light in vacuum, λ is the wavelength of the excitation, k is the Boltzmann constant and T is temperature.

$$k_{TADF} = \frac{2\pi e^2 n^3 f}{\epsilon_0 m_e c \lambda^2} \times \frac{1}{1 + 3\exp(\Delta E_{ST}/kT)} \qquad (1)$$

Experimental calibrations

TD-DFT calculations with many functionals pathologically underestimate the excitation energy of charge-transfer excitations due to the semi-local nature of the density functional theory (DFT) exchange–correlation kernel[35,36]. Interestingly, the amount of exact Hartree–Fock exchange in the B3LYP functional results in cost-effective accuracy for TADF donor–acceptor systems, performing on-par or better than long-range corrected and other hybrid functionals[37–39]. To confirm the accuracy of this approach, we performed a linear calibration scheme using 46 experimental data points from the literature, and verified our calibration against 17 in-house measurements. The calibration is reported in Fig. 2a. It was found that experimental data for the maximum photoluminescent emission wavelength in toluene solution, itself a commonly accepted proxy for electroluminescence wavelength, showed excellent correlation with the vertical absorption calculated with TD-DFT/B3LYP/6-31G(d). We used this calibration scheme to account for the systematic error of DFT, the effects of the molecular environment and the magnitude of the Stokes shift. Inclusion of continuous solvent models in the calculations to account for the effect of the dielectric screening of the host material did not result in appreciable improvement to the calibration. Performance was also not improved with other DFT functionals M06-2X, BH&HLYP, LC-ωPBE0, CAM-B3LYP, and ω-B97X-D or using the polarizable continuum model ($\epsilon = 3$) (the molecules included in the calibration, together with the performance of the different functionals are described in section 4 of the Supplementary Information). Calculations using the equilibrium geometries from the TD-DFT geometry optimizations on the S_1 surface were found to afford inferior results. Despite the fact that the amount of Hartree–Fock exchange in B3LYP results in accurate predictions for vertical absorption, the TD-B3LYP potential energy surface overestimates the charge-transfer character of the S_1 state by predicting very low values of ΔE_{ST} and f. Since all the chosen candidates featured charge-transfer excitations, where the singlet and triplet states have similar electronic density distributions, we used UDFT on the T_1 surface for the geometry optimization and
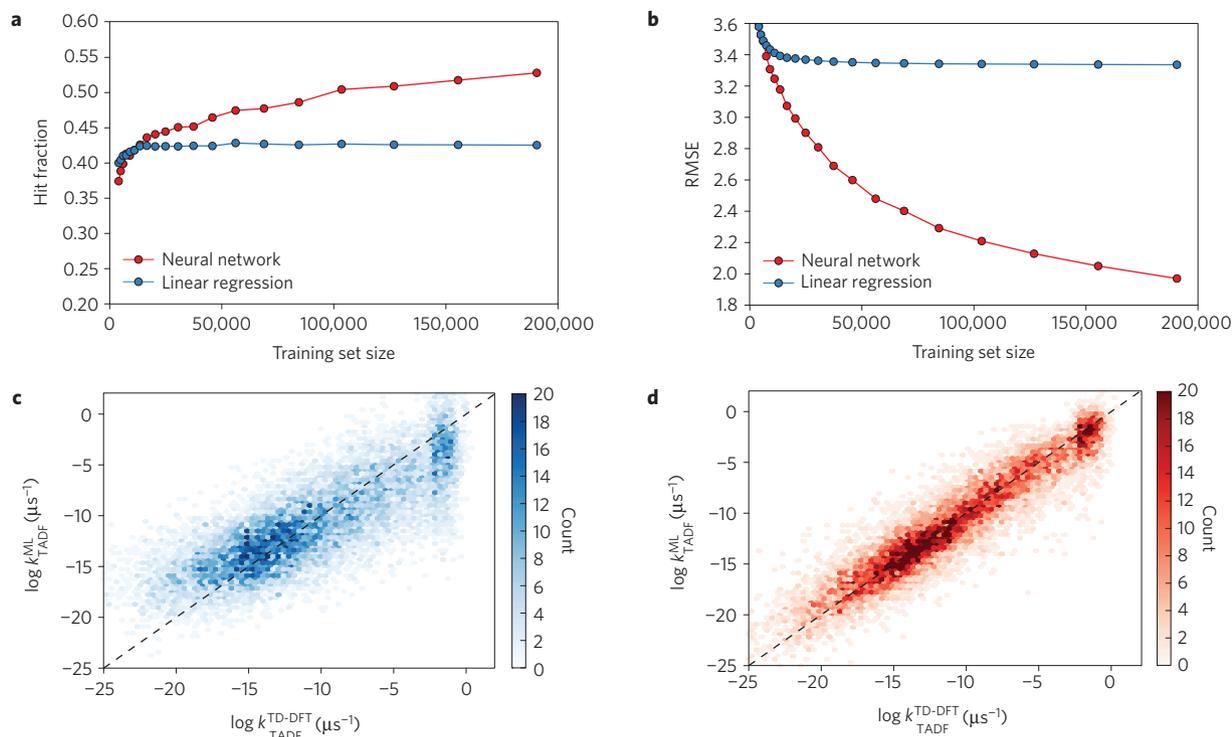
Figure 3 | Effectiveness of machine learning. a, Fraction of molecules in the test set correctly ranked in the top 5%, as a function of the amount of training data. **b**, Root mean square error (RMSE) in $\log(k_{TADF})$ as a function of the amount of training data. **c**, Linear model predictions against the TD-DFT-derived data with the largest training set. $R^2 = 0.80$. **d**, Neural network predictions against the TD-DFT-derived data with the largest training set. $R^2 = 0.94$.

obtained slightly superior performance to that of the S_0 minima, but at an increased computational cost.

Empirical ΔE_{ST} values have significant experimental uncertainty. The three most commonly used methods to measure ΔE_{ST} are peak and onset spectroscopic estimates of the S_1 and T_1 energies and thermal activation energies extracted from Arrhenius plots. These approaches are known to report widely varying estimates of ΔE_{ST} for the same compounds. Spectroscopic estimates have added difficulty in extracting accurate energies from the broad emission spectra of charge-transfer compounds.

TD-DFT calculations afford reasonably accurate results for the Arrhenius activation energy, yielding a linear fit with coefficient of determination (R^2) of 0.84 and root mean square error (RMSE) of 0.08 eV, (Fig. 2b), but perform less well for the spectroscopic estimates. Favourable error cancellation arises in particular when both the singlet and the triplet states involve the same orbital transitions, as is often the case in TADF emitters. However, for large-distance charge-transfer excitations, the incorrect asymptotic behaviour of DFT exchange results in the underestimation of ΔE_{ST}. This statistical uncertainty is of particular relevance because k_{TADF} depends exponentially on ΔE_{ST}. Since ΔE_{ST} is limited to positive values in TADF emitters, the random error for low predicted ΔE_{ST} values is one-sided, with a floor at $\Delta E_{ST} = 0$.

Machine learning for pre-screening

At over 1.6 million candidates, the pool of potential molecules was too large to be screened exhaustively with quantum simulation. It was, however, possible to use results from previous calculations to predict which molecules were likely to have good outcomes, and prioritize the molecules to be simulated next.

An empirical model of TD-DFT simulations was built based on the molecules that had already been screened[40,41]. This model estimates a function whose input is a molecule, and whose output is a prediction for k_{TADF}. We used a neural network to create this model[42]. First, each molecule was converted

from a simplified molecular-input line-entry system (SMILES) representation to a fixed-dimensional vector using Extended-Connectivity Fingerprints (ECFP)[43]. The network had two hidden layers with 100 rectified linear (ReLu)[44] units each. The network was regularized using dropout[45]. The number of hidden units, the learning rate, and momentum decay were determined using 100 iterations of Bayesian optimization[46]. The network was trained to minimize the RMSE of predicted $\log(k_{TADF})$ using the Autograd automatic differentiation package[47].

Figure 3a,b compares the accuracy of linear regression against the accuracy of the neural network on a validation data set, as the amount of training data increases. Figure 3c,d shows predicted versus quantum-chemically determined k_{TADF} for both models. The training data used for these plots was chosen randomly from the candidate pool.

During screening, the molecules ranked highest according to neural network predictions were promoted to TD-DFT simulation. The neural network was periodically re-trained as new calculations were added to the data set, to improve accuracy. Because no training data was available initially, the first 40,000 molecules screened were chosen randomly from the candidate library.

Analysis and lead discovery

The results from the quantum calculations exposed thousands of emitters expected to be highly efficient, with about 900 (0.25% of the screened molecules) being extremely promising, having $f > 0.1$ and $\Delta E_{ST} < 0.15$ eV, and approximately 3,000 with $f > 0.05$ and $\Delta E_{ST} < 0.175$ eV. These values are expected to be optimal for both the prompt and delayed quantum yields, and result in high EQE if incorporated into an appropriately optimized device. Using a threshold of $f > 0.025$ and $\Delta E_{ST} < 0.15$ eV, the search was able to recover many well-known TADF emitters, such as 4CzIpn, 4CzTPN, Cis-BOx2, Trans-BOx2, PXZ-OXD, 2PXZ-OXD, PXZ-TRZ, m-ATP-CDP, bis-PXZ-TRZ, TXO-TPA or methylated derivatives of DPA-CZ TRZ[14].
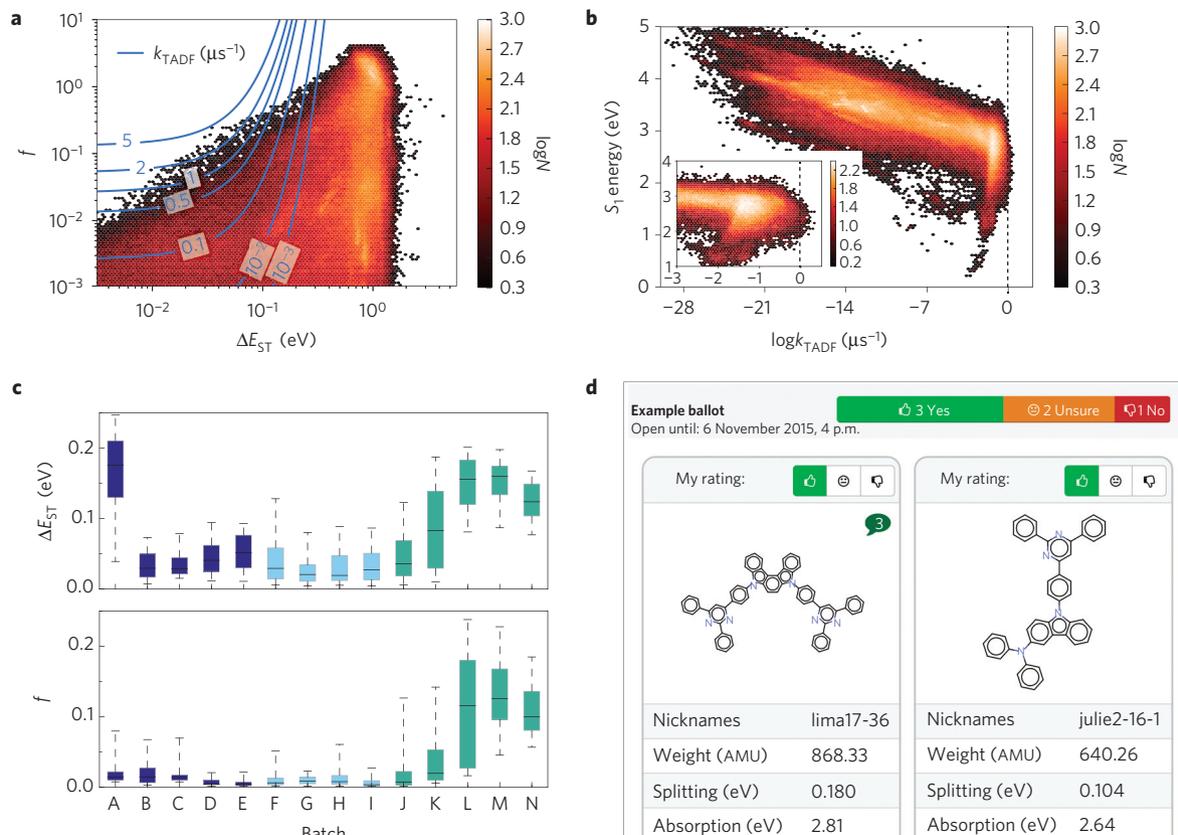
Figure 4 | Candidate statistics and voting tool. a, Number of screened molecules as a function of singlet–triplet splitting (ΔE_{ST}) and oscillator strength (f). Contour lines represent estimated k_{TADF} (μs^{-1}) assuming S_1 at 3.0 eV. **b**, Number of screened molecules as a function of k_{TADF} and S_1 energy. Vertical dashed line corresponds to $k_{TADF} = 1\,\mu s^{-1}$. **c**, ΔE_{ST} and f for decision batches A–N, colour-coded by each of the generations in the candidate library. The box-and-whisker plots represent the statistical distribution of predictions for all the candidates in each batch. The bottom and top of the bar are the first and third quartiles, and the band inside the box is the median. The lines extending vertically from the boxes indicate the maximum and minimum of the range. **d**, Screenshot from the interactive web tool for molecular voting.

Figure 4a,b shows aggregate statistics for the predicted optical properties of over 400,000 emitters. The well-known tradeoff between f and ΔE_{ST}, whose overcoming is at the core of current molecular design efforts,[28,48] is evident in Fig. 4a. The contour lines surround the areas where high k_{TADF} character is expected, and thus high quantum yield in the delayed component. Regarding the prompt component, the radiative rate, and thus EQE, increase along the f axis for any given k_{TADF} contour. Figure 4b shows the counts of emitters as a function of k_{TADF} and the S_1 energy. By design, the dyes are expected to have local excitations in the ultraviolet and deep blue, and only afford redshifted states when charge-transfer transitions are available. Our screening approach is highly efficient in generating potential TADF compounds all across the visible spectrum. Both Fig. 4a and b show the low abundance of molecules with high k_{TADF} and the rapid fall as k_{TADF} reaches $1\,\mu s^{-1}$. Our theoretical predictions for a combined data set of over 1.6 million molecules suggest that the delayed decay rates in TADF emitters have an upper bound above the $1\,\mu s^{-1}$ decay rate at which phosphorescent materials plateau, underscoring the competitiveness of TADF emitters with phosphors. In addition, the Pareto frontiers in Fig. 4a,b demonstrate an apparent fundamental limit to the decay rate. This underscores the potential of HTVS to explore the intrinsic boundaries of chemical space.

Collaborative decision-making
After computation, the search space was reduced to human-tractable decision batches using a web-based selection process that included data visualization and sorting interfaces. The predicted properties of the batches are summarized in Fig. 4c. Each batch contained around one hundred lead candidates that shared some structural features and represented the most promising candidates screened thus far regarding predicted optical properties. The total number of candidates analysed through batches nears 2,500. Human experts assessed the candidates in each cycle using custom web voting tools (Fig. 4d). Voters were asked to consider predicted properties, novelty and synthetic accessibility. A group of two to six synthetic organic chemists rated the molecules on a positive/neutral/negative scale. Finally, a small consensus set of molecules from the batch was selected for synthesis and characterization in devices. The web tools also provided a catalyst for group discussion. The effort can be quantified around 4 h per person per batch for this decision-making step.

Experimental validation
Selected candidates from the search described above were synthesized and tested experimentally to assess the predictive power of the screening process (Fig. 5a). The molecules contain phenoxazine, indolocarbazole, and 3-diphenylaminocarbazole donors combined with pyridine and pyrimidine acceptors.

The candidates were used to explore the balance between pure fluorescence (high f, high ΔE_{ST}) and pure TADF (low f, low ΔE_{ST}) with emphasis in the high-performing intermediate regime where both can be leveraged ($\Delta E_{ST} \sim 0.1$–0.2 eV, $f > 0.1$). Table 1 reports the predicted electronic properties of the four emitters and experimental measurements in solution, thin films and optoelectronic devices. Section 1.2 of the Supplementary
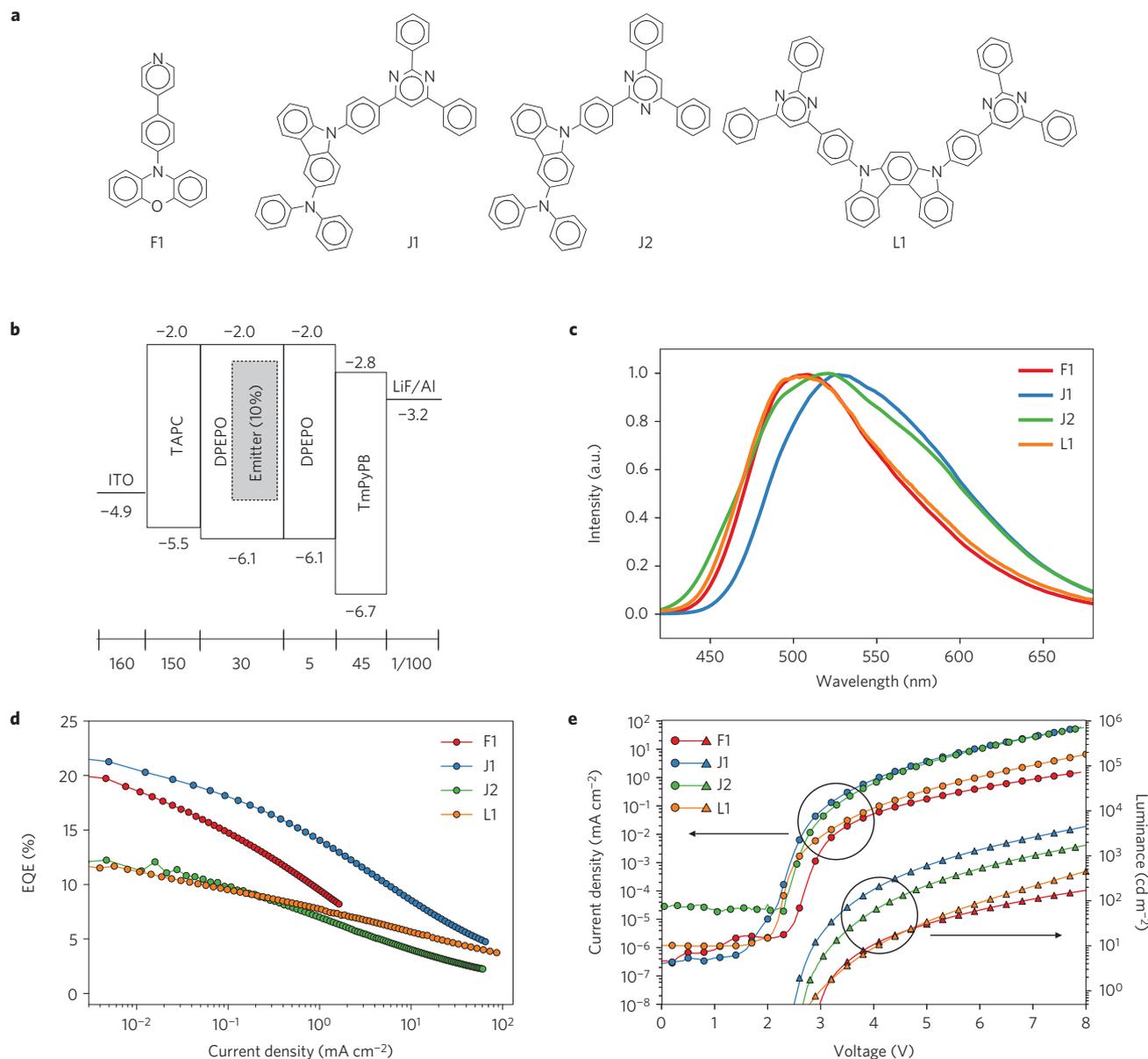
Figure 5 | Lead candidates and optoelectronic characterization. a, Promising molecules that were synthesized and tested. Compound abbreviations are composed by the first letter of the batch of origin and a running index. **b**, Device structure and energy band diagram of lead candidates. Energies in eV. Thickness in nm. **c**, Electroluminescence spectra. **d**, External quantum efficiency as a function of current density. **e**, Current density and luminance as a function of applied voltage.

Information provides a detailed description of experimental procedures and the spectroscopic results. In general, our theoretical predictions agree with experiments within the known accuracy of TD-DFT and the noise in experimental measurements, which is particularly significant for ΔE_{ST}. The estimated values for emission wavelength in toluene, ΔE_{ST}, f and k_{TADF} all show to be predictive in the context of an HTVS effort, with mean unsigned errors of 7 nm, 0.1 eV, 0.05 and $0.1\,\mu s^{-1}$, respectively. TD-DFT calculations at the T_1 (UDFT) and S_1 (TD-DFT) energy surfaces minima are reported in section 4 of the Supplementary Information. The T_1 approach results in qualitatively similar results, whereas the S_1 calculations largely overestimate the charge-transfer character of the four emitters.

For ease of comparison, a common device architecture (Fig. 5b) was used for all candidates, based on initial tests of J1. This leaves room for optimization around any given candidate. Electroluminescence spectra are reported in Fig. 5c, where the broad

features of charge-transfer excited states that are characteristic in TADF are evident. Figure 5d shows the EQE as a function of current density, and L1 is observed to have a less severe roll-off than the other emitters, which is suggestive of higher chemical and operational stability. Figure 5 reports the electrical and electroluminescent properties of the devices as a function of applied voltage.

F1 was chosen for its high predicted k_{TADF} value and very low predicted f and ΔE_{ST}; thus, it would be expected to display high TADF character. These predictions are in keeping with the experimental ΔE_{ST} and f values and also with the low observed photoluminescence quantum yield (PLQY). The maximum EQE of F1 is over 20%, despite the low fluorescence PLQY, confirming an extreme TADF character and very efficient RISC in the device. J1, J2 and L1 explore the balance and tradeoffs between prompt and delayed components. J1 and J2 also highlight the subtlety of structure–property relationships in TADF and the ability of

Table 1 | Selected molecules and devices tested.

	Theory				Experiment												
					ΔE_{ST} Tol. (meV)		PL_{max} (nm)*		PLQY Tol. (%)		f						
Cpd	PL (nm)	ΔE_{ST} (meV)	f	k_{TADF} (μs^{-1})	Onset	Max	Tol.	Thin film	Air	N_2	Abs†	Em‡	τ_p (ns)	τ_d (μs) [k_d]	EL_{max} (nm)§	Max EQE
F1	485	15	0.003	0.141	100	−160	486	456‖	10	13	0.020	0.016	6.6	6.1 [0.16]	507	20.6
J1	485	104	0.124	0.221	60	−140	490	503¶	42	67	0.153	0.060	7.4	2.2 [0.45]	524	22.0
J2	485	36	0.001	0.004	180	0	462	502¶	24	40	0.051	0.025	10	9.2 [0.11]	520	12.7
L1	451	179	0.257	0.029	320	110	451	494¶	48	67	0.107	0.040	11	9.5 [0.10]	501	11.9

*Maximum wavelength of photoluminescence. †Determined from absorption spectra in toluene solution (details in section 1.2 of the Supplementary Information). ‡Determined from prompt rate in DPEPO films (details in section 1.2 of the Supplementary Information). §Maximum wavelength of electroluminescence. ‖Drop-cast thin film. ¶Co-evaporation in DPEPO. $\tau_p = \tau_{prompt}$; $\tau_d = \tau_{delay}$; $k_d = k_{delay}$.

TD-DFT screening to detect them: the position of one heteroatom on the pyrimidine ring is enough to shift between regimes. J1 displays large f, high prompt PLQY and low measured ΔE_{ST}, and it is highly efficient in the device, with a maximum EQE of 22% combining both fluorescent and RISC yield. J2 also has significant TADF character, but its poor fluorescent character (low f and low PLQY) lowers both the prompt and delayed yield, affording a lower maximum EQE of around 12%. L1 shows high PLQY and large f, suggesting efficient prompt fluorescence, but the larger ΔE_{ST} results in lower efficiency in RISC.

Conclusions

We have performed an integrated computation-driven search targeting novel TADF OLED emitters. The discovery process has explored over a million candidates by combining hundreds of thousands of quantum simulations, machine learning, and domain expertise from synthetic chemists, device scientists and industry partners, and ultimately the experimental realization of lead candidates. Excellent predictive power has resulted in the report of devices with over 22% EQE. Nearly one thousand of the identified molecules across the visible spectrum are expected to match or surpass this performance. Detailed predictions for untested leading candidates from the batches A–N are reported in section 2 of the Supplementary Information. Our large data set thoroughly charts the intrinsic limitations of TADF emitters, suggesting that computational exploration of chemical space can be leveraged not only to single out promising new molecules, but also to reveal fundamental chemical insight.

In addition to the present realization in the area of TADF emitters, it must be stressed that this approach to molecular discovery and optimization is more universal. Its modular and collaborative nature make it applicable both in the area of organic electronics and beyond, to fields such as catalysis, high-performance materials, or polymers, where predictive theory, state-of-the-art computer science and accumulated chemical intuition can be combined to expedite discovery.

Methods

Methods and any associated references are available in the online version of the paper.

References

1. Polishchuk, P. G., Madzhidov, T. I. & Varnek, A. Estimation of the size of drug-like chemical space based on GDB-17 data. *J. Comput. Aided Mol. Des.* **27,** 675–679 (2013).
2. Shoichet, B. K. Virtual screening of chemical libraries. *Nature* **432,** 862–865 (2004).
3. Curtarolo, S. *et al.* The high-throughput highway to computational materials design. *Nature Mater.* **12,** 191–201 (2013).
4. Pyzer-Knapp, E. O., Suh, C., Gómez-Bombarelli, R., Aguilera-Iparraguirre, J. & Aspuru-Guzik, A. What is high-throughput virtual screening? A perspective from organic materials discovery. *Annu. Rev. Mater. Res.* **45,** 195–216 (2015).
5. Yang, K., Setyawan, W., Wang, S., Buongiorno Nardelli, M. & Curtarolo, S. A search model for topological insulators with high-throughput robustness descriptors. *Nature Mater.* **11,** 614–619 (2012).
6. Carrete, J., Li, W., Mingo, N., Wang, S. & Curtarolo, S. Finding unprecedentedly low-thermal-conductivity half-Heusler semiconductors via high-throughput materials modeling. *Phys. Rev. X* **4,** 011019 (2014).
7. Huskinson, B. *et al.* A metal-free organic–inorganic aqueous flow battery. *Nature* **505,** 195–198 (2014).
8. Er, S., Suh, C., Marshak, M. P. & Aspuru-Guzik, A. Computational design of molecules for an all-quinone redox flow battery. *Chem. Sci.* **6,** 885–893 (2015).
9. Hachmann, J. *et al.* The Harvard Clean Energy Project: large-scale computational screening and design of organic photovoltaics on the world community grid. *J. Phys. Chem. Lett.* **2,** 2241–2251 (2011).
10. Shin, Y., Liu, J., Quigley, J. J., Luo, H. & Lin, X. Combinatorial design of copolymer donor materials for bulk heterojunction solar cells. *ACS Nano* **8,** 6089–6096 (2014).
11. Hachmann, J. *et al.* Lead candidates for high-performance organic photovoltaics from high-throughput quantum chemistry—the Harvard Clean Energy Project. *Energy Environ. Sci.* **7,** 698–704 (2014).
12. Baldo, M. A. *et al.* Highly efficient phosphorescent emission from organic electroluminescent devices. *Nature* **395,** 151–154 (1998).
13. Yersin, H. *Transition Metal and Rare Earth Compounds* 1–26 (Springer, 2004).
14. Jou, J.-H., Kumar, S., Agrawal, A., Li, T.-H. & Sahoo, S. Approaches for fabricating high efficiency organic light emitting diodes. *J. Mater. Chem. C* **3,** 2974–3002 (2015).
15. Tao, Y. *et al.* Thermally activated delayed fluorescence materials towards the breakthrough of organoelectronics. *Adv. Mater.* **26,** 7931–7958 (2014).
16. Endo, A. *et al.* Efficient up-conversion of triplet excitons into a singlet state and its application for organic light emitting diodes. *Appl. Phys. Lett.* **98,** 083302 (2011).
17. Zhang, Q. *et al.* Design of efficient thermally activated delayed fluorescence materials for pure blue organic light emitting diodes. *J. Am. Chem. Soc.* **134,** 14706–14709 (2012).
18. Zhang, Q. *et al.* Efficient blue organic light-emitting diodes employing thermally activated delayed fluorescence. *Nature Photon.* **8,** 326–332 (2014).
19. Yersin, H., Rausch, A. F., Czerwieniec, R., Hofbeck, T. & Fischer, T. The triplet state of organo-transition metal compounds. Triplet harvesting and singlet harvesting for efficient OLEDs. *Coord. Chem. Rev.* **255,** 2622–2652 (2011).
20. Dias, F. B. *et al.* Triplet harvesting with 100% efficiency by way of thermally activated delayed fluorescence in charge transfer OLED emitters. *Adv. Mater.* **25,** 3707–3714 (2013).
21. Jankus, V. *et al.* Highly efficient TADF OLEDs: How the emitter–host interaction controls both the excited state species and electrical properties of the devices to achieve near 100% triplet harvesting and high efficiency. *Adv. Funct. Mater.* **24,** 6178–6186 (2014).
22. Zhang, Q. *et al.* Nearly 100% internal quantum efficiency in undoped electroluminescent devices employing pure organic emitters. *Adv. Mater.* **27,** 2096–2100 (2015).
23. Tanaka, H., Shizu, K., Lee, J. & Adachi, C. Effect of atom substitution in chalcogenodiazole-containing thermally activated delayed fluorescence emitters on radiationless transition. *J. Phys. Chem. C* **119,** 2948–2955 (2015).
24. Lee, J. *et al.* Controlled emission colors and singlet–triplet energy gaps of dihydrophenazine-based thermally activated delayed fluorescence emitters. *J. Mater. Chem. C* **3,** 2175–2181 (2015).
25. Wang, H. *et al.* Novel thermally activated delayed fluorescence materials–thioxanthone derivatives and their applications for highly efficient OLEDs. *Adv. Mater.* **26,** 5198–5204 (2014).

26. Lee, D. R., Hwang, S.-H., Jeon, S. K., Lee, C. W. & Lee, J. Y. Benzofurocarbazole and benzothienocarbazole as donors for improved quantum efficiency in blue thermally activated delayed fluorescent devices. *Chem. Commun.* **51,** 8105–8107 (2015).
27. Shizu, K. *et al.* Strategy for designing electron donors for thermally activated delayed fluorescence emitters. *J. Phys. Chem. C* **119,** 1291–1297 (2015).
28. Sagara, Y. *et al.* Highly efficient thermally activated delayed fluorescence emitters with a small singlet–triplet energy gap and large oscillator strength. *Chem. Lett.* **44,** 360–362 (2015).
29. Shu, Y. & Levine, B. G. Simulated evolution of fluorophores for light emitting diodes. *J. Chem. Phys.* **142,** 104104 (2015).
30. Zhang, X. *et al.* Theoretical investigation of dihydroacridine and diphenylsulphone derivatives as thermally activated delayed fluorescence emitters for organic light-emitting diodes. *RSC Adv.* **5,** 51586–51591 (2015).
31. RDKit: open source cheminformatics software http://www.rdkit.org (accessed 23 June 2015).
32. Ertl, P. & Schuffenhauer, A. Estimation of synthetic accessibility score of drug-like molecules based on molecular complexity and fragment contributions. *J. Cheminform.* **1,** 8 (2009).
33. Baleizão, C. & Berberan-Santos, M. N. Thermally activated delayed fluorescence as a cycling process between excited singlet and triplet states: application to the fullerenes. *J. Chem. Phys.* **126,** 204510 (2007).
34. Hilborn, R. C. Einstein coefficients, cross sections, *f* values, dipole moments, and all that. *Am. J. Phys.* **50,** 982–986 (1982).
35. Dreuw, A. & Head-Gordon, M. Single-reference *ab initio* methods for the calculation of excited states of large molecules. *Chem. Rev.* **105,** 4009–4037 (2005).
36. Gritsenko, O. & Baerends, E. J. Asymptotic correction of the exchange–correlation kernel of time-dependent density functional theory for long-range charge-transfer excitations. *J. Chem. Phys.* **121,** 655–660 (2004).
37. Huang, S. *et al.* Computational prediction for singlet- and triplet-transition energies of charge-transfer compounds. *J. Chem. Theory Comput.* **9,** 3872–3877 (2013).
38. Penfold, T. J. On predicting the excited-state properties of thermally activated delayed fluorescence emitters. *J. Phys. Chem. C* **119,** 13535–13544 (2015).
39. Moral, M., Muccioli, L., Son, W-J., Olivier, Y. & Sancho-García, J. C. Theoretical rationalization of the singlet–triplet gap in OLEDs materials: impact of charge-transfer character. *J. Chem. Theory Comput.* **11,** 168–177 (2015).
40. Bergeron, C., Krein, M., Moore, G., Breneman, C. M. & Bennett, K. P. Modeling choices for virtual screening hit identification. *Mol. Inform.* **30,** 765–777 (2011).
41. Eckert, H. & Bajorath, J. Molecular similarity analysis in virtual screening: foundations, limitations and novel approaches. *Drug Discov. Today* **12,** 225–233 (2007).
42. Dahl, G. E., Jaitly, N. & Salakhutdinov, R. Multi-task neural networks for QSAR predictions. Preprint at http://arXiv.org/abs/1406.1231 (2014).
43. Rogers, D. & Hahn, M. Extended-connectivity fingerprints. *J. Chem. Inf. Model.* **50,** 742–754 (2010).
44. Nair, V. & Hinton, G. E. Rectified linear units improve restricted Boltzmann machines. *Proc. 27th International Conference on Machine Learning (ICML-10)* 807–814 (2010).
45. Srivastava, N., Hinton, G., Krizhevsky, A., Sutskever, I. & Salakhutdinov, R. Dropout: a simple way to prevent neural networks from overfitting. *J. Mach. Learn. Res.* **15,** 1929–1958 (2014).
46. Snoek, J., Larochelle, H. & Adams, R. P. Practical Bayesian optimization of machine learning algorithms. *Adv. Neural Inf. Process. Syst.* **25,** 2951–2959 (2012).
47. Maclaurin, D., Duvenaud, D. & Johnson, M. J. HIPS/autograd GitHub https://github.com/HIPS/autograd (accessed 29 October 2015).
48. Hirata, S. *et al.* Highly efficient blue electroluminescence based on thermally activated delayed fluorescence. *Nature Mater.* **14,** 330–336 (2015).

Acknowledgements

Author contributions

A.A.-G., M.B. and R.P.A. conceived the project. T.D.H. designed and wrote the custom computer code for molecular screening, with contributions from R.G.-B. and J.A.-I. R.G.-B. and J.A.-I. designed the molecules, with contributions from A.A.-G., H.M., M.N. and H.S.C. R.G.-B. and J.A.-I. performed calculations and analysed theoretical predictions. M.A.B.-F. carried out the experimental calibration of the theoretical methods. D.M., D.D. and R.P.A. applied machine learning to the computational predictions. H.S.C. and G.M. assessed synthetic feasibility of molecular candidates, with contributions from W.H., S.J., H.M., M.N. and S.K. R.G.-B., J.A.-I., T.D.H., H.S.C., M.A.B.-F., G.M., D.M., D.D., S.H., S.J., H.M., M.N., S.K., R.P.A., M.B. and A.A.-G. selected the molecules for characterization. S.J. synthesized J1-2 and L1. S.J., H.S.C., T.W., D.-G.H. and M.E. collected and analysed spectroscopic data. D.-G.H., M.E. and T.W. manufactured and tested devices for F1, J1, J2 and L1, with contributions from H.K. R.G.-B., J.A.-I. and T.D.H. wrote the first version of the manuscript. All authors contributed to the discussion, writing and editing of the manuscript. A.A.-G. and R.P.A. supervised the computational chemistry study. R.P.A. and A.A.-G. supervised the machine learning approach. M.B. supervised the device fabrication.

Additional information

Supplementary information is available in the online version of the paper. Reprints and permissions information is available online at www.nature.com/reprints. Correspondence and requests for materials should be addressed to A.A.-G.

Competing financial interests

The authors declare no competing financial interests.

Methods

Quantum chemistry. Molecular libraries were enumerated using the RDKit cheminformatics package[31]. Initial 3D conformations were generated at the force-field level (MMFF94) using a random distance-matrix approach as implemented in RDKit[49]. Ground-state geometries were re-optimized using PBE/6-31(d) and then refined at the B3LYP/6-31(d) level. Excited-state calculations were carried out with TD-DFT/B3LYP/6-31(d) using both the Tamm–Dancoff approximation (TDA) and the random phase approximation (RPA). Molecules were screened using the lowest energy force-field conformation, and for the more promising candidates the search was expanded to other conformers. In this case, the Boltzmann averages of DFT properties are reported. Candidate molecules were first assessed using ground-state geometries to estimate both f and ΔE_{ST}, and thus k_{TADF}. Promising leads were further filtered based on ΔE_{ST} using equilibrium geometries on the T_1 surface. Ground-state DFT calculations were performed on CPUs with Q-Chem v4.1.2[50] and GPUs with Terachem v1.5[51]. TD-DFT calculations were performed using Q-Chem v4.1.2. The average total run time for each molecule through the computational tree was 39 core-hours and the grand total of CPU time for the entire HTVS project is around 13 million core-hours.

Chemical synthesis. The procedures followed for the chemical synthesis and characterization of novel compounds is reported in section 1.1 of the Supplementary Information.

Optical and redox characterization. The ultraviolet–visible (UV–Vis) absorption and solution photoluminescence (PL) emission spectra of materials were obtained from dilute toluene solution (1×10^{-5} M), while the solid PL spectra were obtained from thin films prepared by vacuum evaporation. UV–Vis spectra were obtained by means of a Varian model UV–Vis–NIR spectrophotometer 5000 and the fluorescence spectra were measured on a HITACHI F7000 spectrometer for the experiments in solution. Thin films of compounds F1, J1, J2 and L1 doped at 10 vol% in bis[2-(diphenylphosphino)phenyl] ether oxide (DPEPO) with a thickness of 70 nm were prepared by thermal evaporation on quartz substrates. We performed optical measurements on those samples through the substrate using a microscope set-up with a $\lambda = 405$ nm pulsed 31.25 kHz excitation laser (PicoQuant LDH) focused to an approximately 10 µm spot size. Time-resolved PL measurements were taken using an avalanche photodiode single-photon detector (PicoQuant PDM). We used a 450 nm longpass filter to block excitation light. Prompt and delayed lifetimes were then obtained from exponential fits. Triplet energy values of the TADF materials were obtained from the photoluminescence spectra at 77 K using liquid nitrogen. PL quantum yields in solution were determined with a Hamamatsu absolute PL quantum yield spectrometer C11347 (Quantaurus-QY). Energy levels were measured by using cyclic voltammetry (CV). Each material was dissolved in anhydrous dichloromethane with 0.1 M tetrabutylammonium hexafluorophosphate as the electrolyte to measure the oxidation from which the HOMO energy level was estimated. A glassy carbon electrode used as the working electrode, a platinum wire used as a counter electrode, and saturated Ag/AgCl was used a reference electrode. Ferrocene was used as the standard reference. All solutions were purged with nitrogen for 10 min before each experiment.

OLED devices. ITO substrates from Luminescence Technology were cleaned by sonicating them in diluted detergent (Micro-90), DI water, acetone and isopropyl alcohol. Subsequently they were boiled in isopropyl alcohol and treated under oxygen plasma for 6 min. The OLED devices with an active area of 1.6 mm^2 for testing the molecules J1 and J3 are composed of the following thin films: indium tin oxide (ITO, 160 nm), 4,4′-cyclohexylidenebis[N,N-bis(4-methylphenyl) benzenamine] (TAPC, 150 nm), 10% emissive molecules doped in DPEPO (30 nm), DPEPO (5 nm), 1,3,5-tris(3-pyridyl-3-phenyl)benzene (TmPyPb, 45 nm), lithium fluoride (LiF, 1 nm) and aluminium (Al, 100 nm). The organic thin films were deposited on cleaned ITO substrates in a vacuum chamber by thermal evaporation at a base pressure below ~10^{-7} torr, at rates below 0.1 nm s^{-1}. LiF and aluminium were deposited in the same thermal evaporator at base pressures below ~10^{-7} torr, at rates below 0.2 nm s^{-1}. The devices were encapsulated using a cover glass and epoxy in a nitrogen glovebox, where the oxygen and moisture levels were kept below 1 ppm. We obtained the voltage, current and electroluminescence data using a precision semiconductor parameter analyzer (4156C, Agilent) and a silicon photodetector (FDS1010, Thorlabs). The OLED was placed directly on top of the photodetector without any intervening optics and the area of the detector is considerably larger than the active area of the device. Consequently, the measured quantum yields are independent of the angular emission of each OLED and the measurement does not require any correction for wide-angle light not collected by the detector. The electroluminescence spectrum of the device was recorded using a spectrometer (SP2300, Princeton Instruments).

Code availability. With the exception of the open source toolkit for cheminformatics RDKit (http://www.rdkit.org), the computer code used in this work is proprietary. The Terachem (http://www.petachem.com) and QChem (http://www.q-chem.com) packages are commercially available.

References

49. Tosco, P., Stiefl, N. & Landrum, G. Bringing the MMFF force field to the RDKit: implementation and validation. *J. Cheminformat.* **6,** 37 (2014).
50. Shao, Y. *et al*. Advances in molecular quantum chemistry contained in the Q-Chem 4 program package. *Mol. Phys.* **113,** 184–215 (2015).
51. Ufimtsev, I. S. & Martinez, T. J. Quantum chemistry on graphical processing units. 3. Analytical energy gradients, geometry optimization, and first principles molecular dynamics. *J. Chem. Theory Comput.* **5,** 2619–2628 (2009).